PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/00, 15/52

A1

(11) International Publication Number: WO 97/45535

(43) International Publication Date: 4 December 1997 (04.12.97)

(21) International Application Number: PCT/EP97/02566

(22) International Filing Date: 20 May 1997 (20.05.97)

(30) Priority Data:
60/018,335 24 May 1996 (24.05.96) US

(71) Applicant *(for all designated States except US)*: NOVARTIS AG [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GENSCHIK, Pascal [FR/FR]; 9, rue du Hameau, F-67170 Hochstett (FR). FILIPOWICZ, Witold [PL/CH]; Burgstrasse 157, CH-4125 Riehen (CH).

(74) Agent: ROTH, Bernhard, M.; Novartis AG, Patent- und Markenabteilung, Klybeckstrasse 141, CH-4002 Basel (CH).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: RECOMBINANT RNA 3'-TERMINAL PHOSPHATE CYCLASES AND PRODUCTION METHODS THEREOF

(57) Abstract

The invention provides a method for producing RNA 3'-terminal phosphate cyclase comprising expressing a nucleic acid sequence encoding said cyclase, or a derivative thereof in a recombinant host cell. The invention also provides a nucleic acid sequence encoding RNA 3'-terminal phosphate cyclase, or a derivative thereof, as well as vectors and host cells containing such nucleic acids. The RNA 3'-terminal phosphate cyclase of the invention is found to be a conserved enzyme across a wide variety of organisms. The invention accordingly provides RNA 3'-terminal phosphate cyclase polypeptides of eukaryotic as well as prokaryotic origin.

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AL** | Albania | **ES** | Spain | **LS** | Lesotho | **SI** | Slovenia |
| **AM** | Armenia | **FI** | Finland | **LT** | Lithuania | **SK** | Slovakia |
| **AT** | Austria | **FR** | France | **LU** | Luxembourg | **SN** | Senegal |
| **AU** | Australia | **GA** | Gabon | **LV** | Latvia | **SZ** | Swaziland |
| **AZ** | Azerbaijan | **GB** | United Kingdom | **MC** | Monaco | **TD** | Chad |
| **BA** | Bosnia and Herzegovina | **GE** | Georgia | **MD** | Republic of Moldova | **TG** | Togo |
| **BB** | Barbados | **GH** | Ghana | **MG** | Madagascar | **TJ** | Tajikistan |
| **BE** | Belgium | **GN** | Guinea | **MK** | The former Yugoslav Republic of Macedonia | **TM** | Turkmenistan |
| **BF** | Burkina Faso | **GR** | Greece | | | **TR** | Turkey |
| **BG** | Bulgaria | **HU** | Hungary | **ML** | Mali | **TT** | Trinidad and Tobago |
| **BJ** | Benin | **IE** | Ireland | **MN** | Mongolia | **UA** | Ukraine |
| **BR** | Brazil | **IL** | Israel | **MR** | Mauritania | **UG** | Uganda |
| **BY** | Belarus | **IS** | Iceland | **MW** | Malawi | **US** | United States of America |
| **CA** | Canada | **IT** | Italy | **MX** | Mexico | **UZ** | Uzbekistan |
| **CF** | Central African Republic | **JP** | Japan | **NE** | Niger | **VN** | Viet Nam |
| **CG** | Congo | **KE** | Kenya | **NL** | Netherlands | **YU** | Yugoslavia |
| **CH** | Switzerland | **KG** | Kyrgyzstan | **NO** | Norway | **ZW** | Zimbabwe |
| **CI** | Côte d'Ivoire | **KP** | Democratic People's Republic of Korea | **NZ** | New Zealand | | |
| **CM** | Cameroon | | | **PL** | Poland | | |
| **CN** | China | **KR** | Republic of Korea | **PT** | Portugal | | |
| **CU** | Cuba | **KZ** | Kazakstan | **RO** | Romania | | |
| **CZ** | Czech Republic | **LC** | Saint Lucia | **RU** | Russian Federation | | |
| **DE** | Germany | **LI** | Liechtenstein | **SD** | Sudan | | |
| **DK** | Denmark | **LK** | Sri Lanka | **SE** | Sweden | | |
| **EE** | Estonia | **LR** | Liberia | **SG** | Singapore | | |

# RECOMBINANT RNA 3'-TERMINAL PHOSPHATE CYCLASES AND PRODUCTION METHODS THEREOF

The present invention relates to a ubiquitously distributed RNA 3'-terminal phosphate cyclase protein.

RNA 3'-terminal phosphate cyclase catalyses the conversion of 3' phosphate to a 2', 3'-phosphodiester at the 3' end of RNA molecules. The activity has been previously detected and partial purifications thereof have been prepared (reviewed in Filipowicz and Vicente, (1990) Meth. Enzymol. 181, 499-510). Its activity results in the activation of the 3' end of RNA molecules. Although this enzyme is unlikely to be involved directly in tRNA splicing, it is postulated that it may be involved in RNA repair in the nucleus or in the cytoplasm, or in other RNA ligation reactions. Most known RNA ligases require 3'-terminal cyclic phosphate in order to perform the ligation reaction. RNA 3'-terminal phosphate cyclase may also be involved in the cyclisation of the 3'-terminal phosphate of the spliceosomal small nuclear RNA U6.

Although RNA 3'-terminal phosphate cyclase appears to be a useful tool for the study of RNA synthesis and metabolism in the cell, as well as for *in vitro* RNA manipulation, it has hitherto not been available except in the form of an activity purified from HeLa cells and also identified in Xenopus oocytes. We have now cloned and sequenced human RNA 3'-terminal phosphate cyclase. Surprisingly, we have found that one of the open reading frames of unknown function in *E. coli* shares significant sequence similarity (32% identitx, 55% similarity) with the human protein. *E. coli* protein, when expressed and purified from recombinant cells, is shown to have RNA 3'-terminal phosphate cyclase activity. Sequences with significant similarity to RNA 3'-terminal phosphate cyclase isoalted from human and *E. coli* sources are also found in other organisms, suggesting a key role for this enzyme in cellular metabolism.

## Summary of the Invention

There is provided a method for producing RNA 3'-terminal phosphate cyclase comprising expressing the a nucleic acid sequence encoding said cyclase, or a derivative thereof in a recombinant host cell. The invention also provides a nucleic acid sequence encoding RNA 3'-terminal phosphate cyclase, or a derivative thereof, as well as vectors and host cells containing such nucleic acids.

The RNA 3'-terminal phosphate cyclase of the invention is found to be a conserved enzyme across a wide variety of organisms. The invention accodingly provides RNA 3'-terminal phosphate cyclase polypeptides of eukaryotic as well as prokaryotic origin.

Detailed Description of the Invention

The method of the first aspect of the invention preferably comprises the steps of transfecting a host cell with a vector comprising a nucleic acid sequence encoding RNA 3'-terminal phosphate cyclase, culturing the host cell to express RNA 3'-terminal phosphate cyclase and purifying RNA 3'-terminal phosphate cyclase from the cell culture.

RNA 3'-terminal phosphate cyclase polypeptide produced by the method of the invention has homology, as hereinafter defined, with the sequence set forth in SEQ ID No. 2. Accordingly, it should be understood that the term " RNA 3'-terminal phosphate cyclase" as used herin refers to any polypeptide which is homologous to this sequence. It has been determined, and is set forth herein, that RNA 3'-terminal phosphate cyclase is a widely conserved polypeptide. As with many widely conseved polypeptides, a number of amino acid mutations have arisen between organisms whose evolutionary paths have diverged. Many of these mutations, however, are conservative and the polypeptides retain a common overall structure and function. Thus, it has been demonstrated that RNA 3'-terminal phosphate cyclase polypeptides from man and *E. coli* retain common functional activity, as well as close structural homology. As used herein, "homology" means that the entities compared share sufficient characteristics for the skilled person to determine that they are similar in origin and function. Each of the RNA 3'-terminal phosphate cyclase polypeptides disclosed herein is homologous to the others. Preferably, homology is used to refer to sequence identity. Thus, the derivatives of RNA 3'-terminal phosphate cyclase preferably retain substantial sequence identity with the RNA 3'-terminal phosphate cyclase polypeptides disclosed herein.

"Substantial homology", where homology indicates sequence identity, means more than 20% sequence identity, preferably more than 35% sequence identity and most preferably a sequence identity of 30% or more.

Preferred sequences of RNA 3'-terminal phosphate cyclases in accordance with the present invention are set out herein as follows:

SEQ ID No. 2 human

| | |
|---|---|
| SEQ ID No. 4 | S. cerevisiae |
| SEQ ID No. 6 | E. coli |
| SEQ ID No. 8 | S. pombe |
| SEQ ID No. 10 | human, partial sequence of related polypeptide |
| SEQ ID No. 12 | D. melanogaster, partial sequence |

Other sequences homologous to the above, which may be identified by conventional methods as exemplified herein, are included within the scope of the present invention.

Moreover, the invention should be understood to embrace methods for the production of polypeptides which are derivatives of RNA 3'-terminal phosphate cyclase, and share at least one common structural determinant therewith.

"Common structural determinant" means that the derivative in question at least one structural feature of RNA 3'-terminal phosphate cyclase. Structural features include epitopes or antigenic sites that are capable of cross-reacting with antibodies raised against a naturally occurring or denatured RNA 3'-terminal phosphate cyclase polypeptide or fragment thereof, possession of amino acid sequence identity with RNA 3'-terminal phosphate cyclase and features having common a structure/function relationship. Thus RNA 3'-terminal phosphate cyclase as provided by the present invention includes amino acid mutants, glycosylation variants and other covalent derivatives of RNA 3'-terminal phosphate cyclase which retain the physiological and/or physical properties of RNA 3'-terminal phosphate cyclase. Further included are naturally occurring variants of RNA 3'-terminal phosphate cyclase found with a particular species, preferably a mammal. Such a variant may be encoded by a related gene of the same gene family, by an allelic variant of a particular gene, or represent an alternative splicing variant of the RNA 3'-terminal phosphate cyclase gene.

Moreover, the invention comprises derivatives of RNA 3'-terminal phosphate cyclase which have been modified by chemical means. Such derivatives include molecules wherein the protein of the invention is covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid. Such a moiety may be a detectable moiety such as an enzyme or a radioisotope.

Derivatives can be fragments of RNA 3'-terminal phosphate cyclase. Fragments of RNA 3'-terminal phosphate cyclase comprise individual domains thereof, as well as smaller polypeptides derived from the domains. Preferably, smaller polypeptides derived from RNA 3'-terminal phosphate cyclase according to the invention define a single feature which is

characteristic of RNA 3'-terminal phosphate cyclase. Fragments may in theory be almost any size, as long as they retain one feature of RNA 3'-terminal phosphate cyclase. Preferably, fragments will be between 5 and 200 amino acids in length. Longer fragments are regarded as truncations of the full-length RNA 3'-terminal phosphate cyclase and generally encompassed by the term "RNA 3'-terminal phosphate cyclase".

Derivatives of RNA 3'-terminal phosphate cyclase also comprise mutants thereof, which may contain amino acid deletions, additions or substitutions, subject to the requirement to maintain at least one feature characteristic of RNA 3'-terminal phosphate cyclase. Thus, conservative amino acid substitutions may be made substantially without altering the nature of RNA 3'-terminal phosphate cyclase, as may truncations from the 5' or 3' ends. Deletions and substitutions may moreover be made to the fragments of RNA 3'-terminal phosphate cyclase comprised by the invention. RNA 3'-terminal phosphate cyclase mutants may be produced from a DNA encoding RNA 3'-terminal phosphate cyclase which has been subjected to *in vitro* mutagenesis resulting e.g. in an addition, exchange and/or deletion of one or more amino acids. For example, substitutional, deletional or insertional variants of RNA 3'-terminal phosphate cyclase can be prepared by recombinant methods and screened for immuno-crossreactivity with the native forms of RNA 3'-terminal phosphate cyclase.

Mutations may be performed by any method known to those of skill in the art. Preferred, however, is site-directed mutagenesis of a nucleic acid sequence encoding the cyclase of interest. A number of methods for site-directed mutagenesis are known in the art, from methods employing single-stranded phage such as M13 to PCR-based techniques (see "PCR Protocols: A guide to methods and applications", M.A. Innis, D.H. Gelfand, J.J. Sninsky, T.J. White (eds.). Academic Press, New York. 1990). Preferably, the commercially available Altered Site II Mutagenesis System (Promega) may be employed, according to the directions given by the manufacturer.

In a second aspect of the present invention, there are provided RNA 3'-terminal phosphate cyclase polypeptides in essentially pure form. It has been determined that RNA 3'-terminal phosphate cyclase is highly conserved throughout all organisms which have been tested. RNA 3'-terminal phosphate cyclase highly homologous to the cloned human HeLa cell RNA 3'-terminal phosphate cyclase set forth in SEQ ID No. 2 has been found in *E. coli*, *D. melanogaster*, *S. pombe* and *S. cervisiae*, as set out hereinbefore. Moreover, ESTs homologous to 3' RNA terminal phosphate cyclase have been found in Zebra fish and *Arabidopsis*. The invention therefore provides RNA 3'-terminal phosphate cyclase

polypeptides and derivatives thereof from any source, where "derivatives" are as hereinbefore defined.

A particularly surprising aspect of the invention is the provision of RNA 3'-terminal phosphate cyclase from prokaryotic sources. *E. coli* RNA 3'-terminal phosphate cyclase is encoded by a coding sequence which, due to a sequencing error, has previously been classified as two separate ORFs. Re-sequencing of the relevant part of the *E. coli* genome demonstrates that the sequence previously believed to comprise two ORFs of unknown function is actually a single ORF encoding an enzyme which is highly homologous to human RNA 3'-terminal phosphate cyclase and is functionally equivalent thereto.

According to a further aspect of the present invention, there is provided a nucleic acid encoding RNA 3'-terminal phosphate cyclase. In addition to being useful for the production of recombinant RNA 3'-terminal phosphate cyclase protein, these nucleic acids are also useful as probes, thus readily enabling those skilled in the art to identify and/or isolate nucleic acid encoding RNA 3'-terminal phosphate cyclase.

Furthermore, nucleic acid according to the invention is useful in a method determining the presence of RNA 3'-terminal phosphate cyclase-specific nucleic acid, said method comprising hybridising the DNA (or RNA) encoding RNA 3'-terminal phosphate cyclase (or complementary thereto) to test sample nucleic acid and determining the presence of RNA 3'-terminal phosphate cyclase. In another aspect, the invention provides nucleic acid sequence that is complementary to, or hybridises under stringent conditions to, a nucleic acid sequence encoding RNA 3'-terminal phosphate cyclase.

The invention also provides a method for amplifying a nucleic acid test sample comprising priming a nucleic acid polymerase (chain) reaction with nucleic acid (DNA or RNA) encoding (or complementary to) RNA 3'-terminal phosphate cyclase.

If required, nucleic acids encoding RNA 3'-terminal phosphate cyclase polypeptides may be cloned from tissues according to established procedures using probes derived from sequences identified herein. In particular, such DNAs can be prepared by:

a) isolating mRNA from suitable cells, selecting the desired mRNA, for example by hybridisation with a DNA probe or by expression in a suitable expression system and screening for expression of the desired polypeptide, preparing single-stranded cDNA complementary to that mRNA, then double-stranded cDNA therefrom, or

b) isolating cDNA from a cDNA library and selecting the desired cDNA, for example using a DNA probe or using a suitable expression system and screening for expression of the desired polypeptide, or

c) incorporating the double-stranded DNA of step a) or b) into an appropriate expression vector,

d) transforming appropriate host cells with the vector and isolating the desired DNA.

Polyadenylated messenger RNA (step a) is isolated by known methods. Isolation methods involve, for example, homogenizing cells in the presence of a detergent and a ribonuclease inhibitor, for example heparin, guanidinium isothiocyanate or mercaptoethanol, extracting the mRNA with a chloroform-phenol mixture, optionally in the presence of salt and buffer solutions, detergents and/or cation chelating agents, and precipitating mRNA from the remaining aqueous, salt-containing phase with ethanol, isopropanol or the like. The isolated mRNA may be further purified by centrifuging in a caesium chloride gradient followed by ethanol precipitation and/or by chromatographic methods, for example affinity chromatography, for example chromatography on oligo(dT) cellulose or on oligo(U) sepharose. Preferably, such purified total mRNA is fractionated according to size by gradient centrifugation, for example in a linear sucrose gradient, or chromatography on suitable size fractionation columns, for example on agarose gels.

The desired mRNA is selected by screening the mRNA directly with a DNA probe, or by translation in suitable cells or cell-free systems and screening the obtained polypeptides.

The selection of the desired mRNA is preferably achieved using a DNA hybridisation probe, thereby avoiding the additional step of translation. Suitable DNA probes are DNAs of known nucleotide sequence consisting of at least 17 nucleotides derived from DNAs encoding RNA 3'-terminal phosphate cyclase.

Synthetic DNA probes are synthesised according to known methods as detailed hereinbelow, preferably by stepwise condensation using the solid phase phosphotriester, phosphite triester or phosphoramidite method, for example the condensation of dinucleotide coupling units by the phosphotriester method. These methods are adapted to the synthesis of mixtures of the desired oligonucleotides by using mixtures of two, three or four nucleotides dA, dC, dG and/or dT in protected form or the corresponding dinucleotide coupling units in the appropriate condensation step as described by Y. Ike et al. (Nucleic Acids Research 11, 477, 1983).

For hybridisation, the DNA probes are labelled, for example radioactively labelled by the well known kinase reaction. The hybridisation of the size-fractionated mRNA with the DNA probes containing a label is performed according to known procedures, i.e. in buffer and salt solutions containing adjuncts, for example calcium chelators, viscosity regulating

compounds, proteins, irrelevant DNA and the like, at temperatures favouring selective hybridisation, for example between 0°C and 80°C, for example between 25°C and 50°C or around 65°C, preferably at around 20° lower than the hybrid double-stranded DNA melting temperature.

Fractionated mRNA may be translated in cells, for example frog oocytes, or in cell-free systems, for example in reticulocyte lysates or wheat germ extracts. The obtained polypeptides are screened for RNA 3'-terminal phosphate cyclase activity or for reaction with antibodies raised against RNA 3'-terminal phosphate cyclase, for example in an immunoassay, for example radioimmunoassay, enzyme immunoassay or immunoassay with fluorescent markers. Such immunoassays and the preparation of polyclonal and monoclonal antibodies are well known in the art and are applied accordingly.

The preparation of a single-stranded complementary DNA (cDNA) from the selected mRNA template is well known in the art, as is the preparation of a double-stranded DNA from a single-stranded DNA. The mRNA template is incubated with a mixture of deoxynucleoside triphosphates, optionally radioactively labelled deoxynucleoside triphosphates (in order to be able to screen the result of the reaction), a primer sequence such as an oligo-dT residue hybridising with the poly(A) tail of the mRNA and a suitable enzyme such as a reverse transcriptase for example from avian myeloblastosis virus (AMV). After degradation of the template mRNA for example by alkaline hydrolysis, the cDNA is incubated with a mixture of deoxynucleoside triphosphates and a suitable enzyme to give a double-stranded DNA. Suitable enzymes are for instance a reverse transcriptase, the Klenow fragment of E. coli DNA polymerase I or T4 DNA polymerase. Usually, a hairpin loop structure formed spontaneously by the single-stranded cDNA acts as a primer for the synthesis of the second strand. This hairpin structure is removed by digestion with S1 nuclease. Alternatively, the 3'-end of the single-stranded DNA is first extended by homopolymeric deoxynucleotide tails prior to the hydrolysis of the mRNA template and the subsequent synthesis of the second cDNA strand.

In the alternative, double-stranded cDNA is isolated from a cDNA library and screened for the desired cDNA (step b). The cDNA library is constructed by isolating mRNA from suitable cells, for example human mononuclear leukocytes or human embryonic epithelial lung cells, and preparing single-stranded and double-stranded cDNA therefrom as described above. This cDNA is digested with suitable restriction endonucleases and incorporated into λ phage, for example λ charon 4A or λ gt11 following established procedures. The cDNA library replicated on nitrocellulose membranes is screened by using

a DNA probe as described hereinbefore, or expressed in a suitable expression system and the obtained polypeptides screened for reaction with an antibody specific for the desired RNA 3'-terminal phosphate cyclase, for example an antibody specific for RNA 3'-terminal phosphate cyclase.

A variety of methods are known in the art for the incorporation of double-stranded cDNA into an appropriate vector (step c). For example, complementary homopolymer tracts may be added to the double-stranded DNA and the vector DNA by incubation in the presence of the corresponding deoxynucleoside triphosphates and an enzyme such as terminal deoxynucleotidyl transferase. The vector and double-stranded DNA are then joined by base pairing between the complementary homopolymeric tails and finally ligated by specific joining enzymes such as ligases. Other possibilities are the addition of synthetic linkers to the termini of the double-stranded DNA, or the incorporation of the double-stranded DNA into the vector by blunt- or staggered-end ligation.

The transformation of appropriate host cells with the obtained hybrid vector (step d) and the selection of transformed host cells (step e) are well known in the art. Hybrid vectors and host cells may be particularly suitable for the production of DNA, or for the production of the desired RNA 3'-terminal phosphate cyclase.

The isolation of the desired DNA is achieved by methods known in the art, for example extraction with phenol and/or chloroform or glass beads. Optionally, the DNA can be further manipulated for example by treatment with mutagenic agents to obtain mutants, or by digestion with restriction enzymes to obtain fragments, modify one or both termini to facilitate incorporation into the vector.

In still another aspect of the invention, the nucleic acid is DNA and further comprises a replicable vector comprising the nucleic acid encoding RNA 3'-terminal phosphate cyclase operably linked to control sequences recognised by a host transformed by the vector. Furthermore the invention provides host cells transformed with such a vector and a method of using a nucleic acid encoding RNA 3'-terminal phosphate cyclase to effect the production of RNA 3'-terminal phosphate cyclase, comprising expressing RNA 3'-terminal phosphate cyclase nucleic acid in a culture of the transformed host cells and, if desired, recovering RNA 3'-terminal phosphate cyclase from the host cell culture.

Nucleic acids according to the invention are preferably provided in isolated form. Isolated RNA 3'-terminal phosphate cyclase nucleic acid includes nucleic acid that is free from at least one contaminant nucleic acid with which it is ordinarily associated in the natural source of RNA 3'-terminal phosphate cyclase nucleic acid or in crude nucleic acid

preparations, such as DNA libraries and the like. Isolated nucleic acid thus is present in other than in the form or setting in which it is found in nature. However, isolated RNA 3'-terminal phosphate cyclase encoding nucleic acid includes RNA 3'-terminal phosphate cyclase nucleic acid in ordinarily RNA 3'-terminal phosphate cyclase-expressing cells where the nucleic acid is in a chromosomal location different from that of natural cells or is otherwise flanked by a different DNA sequence than that found in nature.

The preferred sequence encoding RNA 3'-terminal phosphate cyclase is that having substantially the same nucleotide sequence as the coding sequences in SEQ ID No. 1, 3, 5, 7, 9 or 11, with the nucleic acid having the same sequence as these coding sequences being most preferred. As used herein, nucleotide sequences which are substantially the same share at least about 90 % identity. However, in the case of splice variants having e.g. an additional exon sequence homology may be lower.

As is understood by persons skilled in the art, nucleic acid sequences may be substantially altered, by virtue of the degeneracy of the genetic code, without affecting the sequence of the polypeptide encoded thereby. Such "degenerate" equivalents, therefore, are included within the scope of the present invention. In particular, the invention includes all sequences which encode the polypeptides set forth in SEQ ID No. 2, 4, 6, 8, 10 or 12.

The nucleic acids of the invention, whether used as probes or otherwise, preferably possess substantial homology to the sequences encoding RNA 3'-terminal phosphate cyclase as set forth herein . The terms "substantial" and "homology" are used as hereinbefore defined with reference to the RNA 3'-terminal phosphate cyclase polypeptide.

Preferably, nucleic acids according to the invention are fragments of the RNA 3'-terminal phosphate cyclase-encoding sequence, or derivatives thereof as hereinbefore defined in relation to polypeptides. Fragments of the nucleic acid sequence of a few nucleotides in length, preferably 5 to 150 nucleotides in length, are especially useful as probes.

Exemplary nucleic acids can alternatively be characterised as those nucleotide sequences which encode a RNA 3'-terminal phosphate cyclase protein and hybridise to the DNA sequences set forth herein, or a selected fragment of said DNA sequence. Preferred are such sequences encoding RNA 3'-terminal phosphate cyclase which hybridise to the sequences set foth in SEQ ID No. 1, 3, 5, 7, 9 or 11. Preferably, hybridisation occurs under conditions of medium or high stringency.

Stringency of hybridisation refers to conditions under which polynucleic acids hybrids are stable. Such conditions are evident to those of ordinary skill in the field. As known to

those of skill in the art, the stability of hybrids is reflected in the melting temperature ($T_m$) of the hybrid which decreases approximately 1 to 1.5°C with every 1% decrease in sequence homology. In general, the stability of a hybrid is a function of sodium ion concentration and temperature. Typically, the hybridisation reaction is performed under conditions of higher stringency, followed by washes of varying stringency.

As used herein, high stringency refers to conditions that permit hybridisation of only those nucleic acid sequences that form stable hybrids in 1 M $Na^+$ at 65-68 °C. High stringency conditions can be provided, for example, by hybridisation in an aqueous solution containing 6x SSC, 5x Denhardt's, 1 % SDS (sodium dodecyl sulphate), 0.1 $Na^+$ pyrophosphate and 0.1 mg/ml denatured salmon sperm DNA as non specific competitor. Following hybridisation, high stringency washing may be done in several steps, with a final wash (about 30 min) at the hybridisation temperature in 0.2 - 0.1x SSC, 0.1 % SDS.

Moderate stringency refers to conditions equivalent to hybridisation in the above described solution but at about 60-62°C. In that case the final wash is performed at the hybridisation temperature in 1x SSC, 0.1 % SDS.

Low stringency refers to conditions equivalent to hybridisation in the above described solution at about 50-52°C. In that case, the final wash is performed at the hybridisation temperature in 2x SSC, 0.1 % SDS.

It is understood that these conditions may be adapted and duplicated using a variety of buffers, e.g. formamide-based buffers, and temperatures. Denhardt's solution and SSC are well known to those of skill in the art as are other suitable hybridisation buffers (see, e.g. Sambrook, *et al.*, eds. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York or Ausubel, *et al.*, eds. (1990) Current Protocols in Molecular Biology, John Wiley & Sons, Inc.). Optimal hybridisation conditions have to be determined empirically, as the length and the GC content of the probe also play a role.

Given the guidance provided herein, the nucleic acids of the invention are obtainable according to methods well known in the art. For example, a DNA of the invention is obtainable by chemical synthesis, using polymerase chain reaction (PCR) or by screening a genomic library or a suitable cDNA library prepared from a source believed to possess RNA 3'-terminal phosphate cyclase and to express it at a detectable level.

Chemical methods for synthesis of a nucleic acid of interest are known in the art and include triester, phosphite, phosphoramidite and H-phosphonate methods, PCR and other autoprimer methods as well as oligonucleotide synthesis on solid supports. These methods may be used if the entire nucleic acid sequence of the nucleic acid is known, or the

sequence of the nucleic acid complementary to the coding strand is available. Alternatively, if the target amino acid sequence is known, one may infer potential nucleic acid sequences using known and preferred coding residues for each amino acid residue.

A nucleic acid encoding RNA 3'-terminal phosphate cyclase may be isolated by screening suitable cDNA or genomic libraries under suitable hybridisation conditions with a probe, i.e. a nucleic acid disclosed herein including oligonucleotides derivable from the sequences set forth herein. Suitable libraries are commercially available or can be prepared e.g. from cell lines, tissue samples, and the like.

Libraries are screened with probes or analytical tools designed to identify the gene of interest or the protein encoded by it. For cDNA expression libraries suitable means include monoclonal or polyclonal antibodies that recognise and specifically bind to RNA 3'-terminal phosphate cyclase; oligonucleotides of about 20 to 80 bases in length that encode known or suspected RNA 3'-terminal phosphate cyclase cDNA from the same or different species; and/or complementary or homologous cDNAs or fragments thereof that encode the same or a hybridising gene. Appropriate probes for screening genomic DNA libraries include, but are not limited to oligonucleotides, cDNAs or fragments thereof that encode the same or hybridising DNA; and/or homologous genomic DNAs or fragments thereof.

An alternative means to isolate the gene encoding RNA 3'-terminal phosphate cyclase is to use PCR technology as described e.g. in section 14 of Sambrook et al., 1989. This method requires the use of oligonucleotide probes that will hybridise to RNA 3'-terminal phosphate cyclase nucleic acid.

As used herein, an oligonucleotide probe is e.g. a single-stranded DNA or RNA that has a sequence of nucleotides that includes between 10 and 50, preferably between 15 and 30 and most preferably at least about 20 contiguous bases that are the same as (or the complement of) an equivalent or greater number of contiguous bases set forth in SEQ ID No. 1. The nucleic acid sequences selected as probes should be of sufficient length and sufficiently unambiguous so that false positive results are minimised. The nucleotide sequences are usually based on conserved or highly homologous nucleotide sequences or regions of RNA 3'-terminal phosphate cyclase. The nucleic acids used as probes may be degenerate at one or more positions. The use of degenerate oligonucleotides may be of particular importance where a library is screened from a species in which preferential codon usage in that species is not known.

Preferred regions from which to construct probes include 5' and/or 3' coding sequences, sequences predicted to encode ligand binding sites, and the like. For example,

either the full-length cDNA clone disclosed herein or fragments thereof can be used as probes. Preferably, nucleic acid probes of the invention are labelled with suitable label means for ready detection upon hybridisation. For example, a suitable label means is a radiolabel. The preferred method of labelling a DNA fragment is by incorporating $\alpha^{32P}$ dATP with the Klenow fragment of DNA polymerase in a random priming reaction, as is well known in the art. Oligonucleotides are usually end-labelled with $\gamma^{32P}$-labelled ATP and polynucleotide kinase. However, other methods (e.g. non-radioactive) may also be used to label the fragment or oligonucleotide, including e.g. enzyme labelling, fluorescent labelling with suitable fluorophores and biotinylation.

After screening the library, e.g. with a portion of DNA including substantially the entire RNA 3'-terminal phosphate cyclase-encoding sequence or a suitable oligonucleotide based on a portion of said DNA, positive clones are identified by detecting a hybridisation signal; the identified clones are characterised by restriction enzyme mapping and/or DNA sequence analysis, and then examined, e.g. by comparison with the sequences set forth herein, to ascertain whether they include DNA encoding a complete RNA 3'-terminal phosphate cyclase (i.e., if they include translation initiation and termination codons). If the selected clones are incomplete, they may be used to rescreen the same or a different library to obtain overlapping clones. If the library is genomic, then the overlapping clones may include exons and introns. If the library is a cDNA library, then the overlapping clones will include an open reading frame. In both instances, complete clones may be identified by comparison with the DNAs and deduced amino acid sequences provided herein.

In order to detect any abnormality of endogenous RNA 3'-terminal phosphate cyclase, genetic screening may be carried out using the nucleotide sequences of the invention as hybridisation probes. Also, based on the nucleic acid sequences provided herein antisense-type therapeutic agents may be designed. In particular reference thereto, it is to be noted that antisense oligonucleotides are based on oligonucleotide probes as hereinbefore defined, and included within the definition thereof. Such oligonucleotides, especially but not only when intended for use as antisense therapeutic agents, may comprise modifications to the oligonucleotide, for example by incorporation of un-natural nucleotide analogues and modifications to natural oligonucleotides. For example, the oligonucleotides may encompass an altered backbone, for example in the form of a phosphorothioate, modifications such as 2'-O-Methyl modifications, or may be in the form of peptide nucleic acids.

It is envisaged that the nucleic acid of the invention can be readily modified by nucleotide substitution, nucleotide deletion, nucleotide insertion or inversion of a nucleotide stretch, and any combination thereof. Such mutants can be used e.g. to produce a RNA 3'-terminal phosphate cyclase mutant that has an amino acid sequence differing from the RNA 3'-terminal phosphate cyclase sequences as found in nature. Mutagenesis may be predetermined (site-specific) or random. A mutation which is not a silent mutation must not place sequences out of reading frames and preferably will not create complementary regions that could hybridise to produce secondary mRNA structure such as loops or hairpins.

In accordance with another embodiment of the present invention, there are provided cells containing the above-described nucleic acids. Such host cells such as prokaryotic, yeast and higher eukaryotic cells may be used for replicating DNA and producing RNA 3'-terminal phosphate cyclase. Suitable prokaryotes include eubacteria, such as Gram-negative or Gram-positive organisms, such as *E. coli*, e.g. *E. coli* K-12 strains, DH5a and HB101, or Bacilli, especially *Bacillus subtilis*. Further hosts suitable for RNA 3'-terminal phosphate cyclase encoding vectors include eukaryotic microbes such as filamentous fungi or yeast, e.g. Saccharomyces cerevisiae. Higher eukaryotic cells include insect and vertebrate cells, particularly mammalian cells. In recent years propagation of vertebrate cells in culture (tissue culture) has become a routine procedure. Examples of useful mammalian host cell lines are epithelial or fibroblastic cell lines such as Chinese hamster ovary (CHO) cells, NIH 3T3 cells, HeLa cells or 293T cells. The host cells referred to in this disclosure comprise cells in *in vitro* culture as well as cells that are within a host animal.

DNA may be stably incorporated into cells or may be transiently expressed using methods known in the art. Stably transfected mammalian cells may be prepared by transfecting cells with an expression vector having a selectable marker gene, and growing the transfected cells under conditions selective for cells expressing the marker gene. To prepare transient transfectants, mammalian cells are transfected with a reporter gene to monitor transfection efficiency.

To produce such stably or transiently transfected cells, the cells should be transfected with a sufficient amount of RNA 3'-terminal phosphate cyclase-encoding nucleic acid to form RNA 3'-terminal phosphate cyclase. The precise amounts of DNA encoding RNA 3'-terminal phosphate cyclase may be empirically determined and optimised for a particular cell and assay.

Host cells are transfected or, preferably, transformed with the above-captioned expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. Heterologous DNA may be introduced into host cells by any method known in the art, such as transfection with a vector encoding a heterologous DNA by the calcium phosphate coprecipitation technique or by electroporation. Numerous methods of transfection are known to the skilled worker in the field. Successful transfection is generally recognised when any indication of the operation of this vector occurs in the host cell. Transformation is achieved using standard techniques appropriate to the particular host cells used.

Transfected or transformed cells are cultured using media and culturing methods known in the art, preferably under conditions, whereby RNA 3'-terminal phosphate cyclase encoded by the DNA is expressed. The composition of suitable media is known to those in the art, so that they can be readily prepared. Suitable culturing media are also commercially available.

The cDNA or genomic DNA encoding native or mutant RNA 3'-terminal phosphate cyclase can be incorporated into vectors for further manipulation. As used herein, vector (or plasmid) refers to discrete elements that are used to introduce heterologous DNA into cells for either expression or replication thereof. Selection and use of such vehicles are well within the skill of the artisan. Many vectors are available, and selection of appropriate vector will depend on the intended use of the vector, i.e. whether it is to be used for DNA amplification or for DNA expression, the size of the DNA to be inserted into the vector, and the host cell to be transformed with the vector. Each vector contains various components depending on its function (amplification of DNA or expression of DNA) and the host cell for which it is compatible. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, a transcription termination sequence and a signal sequence. Incorporation of cloned DNA into a suitable expression vector, transfection of cells with a plasmid vector or a combination of plasmid vectors, each encoding one or more distinct genes or with linear DNA, and selection of transfected cells are well known in the art (see, e.g. Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press).

Particularly useful for practising the present invention are expression vectors that provide for the transient expression of DNA encoding RNA 3'-terminal phosphate cyclase in

mammalian cells. Transient expression usually involves the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector, and, in turn, synthesises high levels of RNA 3'-terminal phosphate cyclase. For the purposes of the present invention, transient expression systems are useful e.g. for identifying RNA 3'-terminal phosphate cyclase mutants, to identify potential phosphorylation sites, or to characterise functional domains of the protein.

Construction of vectors according to the invention employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing RNA 3'-terminal phosphate cyclase expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridisation, using an appropriately labelled probe based on a sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

Both expression and cloning vectors generally contain nucleic acid sequence that enable the vector to replicate in one or more selected host cells. Typically in cloning vectors, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria, yeast and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (e.g. SV 40, polyoma, adenovirus) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors unless these are used in mammalian cells competent for high level DNA replication, such as COS cells.

Most expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another organism for expression. For example, a vector is cloned in *E. coli* and then the same vector is transfected into yeast or mammalian cells even though it is not capable of replicating independently of the host cell chromosome. DNA may also be replicated by insertion into the host genome. However,

the recovery of genomic DNA encoding RNA 3'-terminal phosphate cyclase is more complex than that of exogenously replicated vector because restriction enzyme digestion is required to excise RNA 3'-terminal phosphate cyclase DNA. DNA can be amplified by PCR and be directly transfected into the host cells without any replication component.

Advantageously, an expression and cloning vector may contain a selection gene also referred to as selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and other toxins, e.g. ampicillin, neomycin, methotrexate or tetracycline, complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.

As to a selective gene marker appropriate for yeast, any marker gene can be used which facilitates the selection for transformants due to the phenotypic expression of the marker gene. Suitable markers for yeast are, for example, those conferring resistance to antibiotics G418, hygromycin or bleomycin, or provide for prototrophy in an auxotrophic yeast mutant, for example the URA3, LEU2, LYS2, TRP1, or HIS3 gene.

Since the replication of vectors is conveniently done in *E. coli*, an *E. coli* genetic marker and an *E. coli* origin of replication are advantageously included. These can be obtained from *E. coli* plasmids, such as pBR322, Bluescript® vector or a pUC plasmid, e.g. pUC18 or pUC19, which contain both *E. coli* replication origin and *E. coli* genetic marker conferring resistance to antibiotics, such as ampicillin.

Suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up RNA 3'-terminal phosphate cyclase nucleic acid, such as dihydrofolate reductase (DHFR, methotrexate resistance), thymidine kinase, or genes conferring resistance to G418 or hygromycin. The mammalian cell transformants are placed under selection pressure which only those transformants which have taken up and are expressing the marker are uniquely adapted to survive. In the case of a DHFR or glutamine synthase (GS) marker, selection pressure can be imposed by culturing the transformants under conditions in which the pressure is progressively increased, thereby leading to amplification (at its chromosomal integration site) of both the selection gene and the linked DNA that encodes RNA 3'-terminal phosphate cyclase. Amplification is the process by which genes in greater demand for the production of a protein critical for growth, together with closely associated genes which may encode a desired protein, are reiterated

in tandem within the chromosomes of recombinant cells. Increased quantities of desired protein are usually synthesised from thus amplified DNA.

Expression and cloning vectors usually contain a promoter that is recognised by the host organism and is operably linked to RNA 3'-terminal phosphate cyclase nucleic acid. Such a promoter may be inducible or constitutive. The promoters are operably linked to DNA encoding RNA 3'-terminal phosphate cyclase by removing the promoter from the source DNA by restriction enzyme digestion and inserting the isolated promoter sequence into the vector. Both the native RNA 3'-terminal phosphate cyclase promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of RNA 3'-terminal phosphate cyclase DNA.

Promoters suitable for use with prokaryotic hosts include, for example, the β-lactamase and lactose promoter systems, alkaline phosphatase, the tryptophan (trp) promoter system and hybrid promoters such as the tac promoter. Their nucleotide sequences have been published, thereby enabling the skilled worker operably to ligate them to DNA encoding RNA 3'-terminal phosphate cyclase, using linkers or adaptors to supply any required restriction sites. Promoters for use in bacterial systems will also generally contain a Shine-Delgarno sequence operably linked to the DNA encoding RNA 3'-terminal phosphate cyclase.

Moreover, the RNA 3'-terminal phosphate cyclase gene according to the invention preferably includes a secretion sequence in order to facilitate secretion of the polypeptide from bacterial hosts, such that it will be produced as a soluble native peptide rather than in an inclusion body. The peptide may be recovered from the bacterial periplasmic space, or the culture medium, as appropriate.

Suitable promoting sequences for use with yeast hosts may be regulated or constitutive and are preferably derived from a highly expressed yeast gene, especially a Saccharomyces cerevisiae gene. Thus, the promoter of the TRP1 gene, the ADHI or ADHII gene, the acid phosphatase (PH05) gene, a promoter of the yeast mating pheromone genes coding for the α- or a-factor or a promoter derived from a gene encoding a glycolytic enzyme such as the promoter of the enolase, glyceraldehyde-3-phosphate dehydrogenase (GAP), 3-phospho glycerate kinase (PGK), hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triose phosphate isomerase, phosphoglucose isomerase or glucokinase genes, or a promoter from the TATA binding protein (TBP) gene can be used. Furthermore,

it is possible to use hybrid promoters comprising upstream activation sequences (UAS) of one yeast gene and downstream promoter elements including a functional TATA box of another yeast gene, for example a hybrid promoter including the UAS(s) of the yeast PH05 gene and downstream promoter elements including a functional TATA box of the yeast GAP gene (PH05-GAP hybrid promoter). A suitable constitutive PHO5 promoter is e.g. a shortened acid phosphatase PH05 promoter devoid of the upstream regulatory elements (UAS) such as the PH05 (-173) promoter element starting at nucleotide -173 and ending at nucleotide -9 of the PH05 gene.

RNA 3'-terminal phosphate cyclase gene transcription from vectors in mammalian hosts may be controlled by promoters derived from the genomes of viruses such as polyoma virus, adenovirus, fowlpox virus, bovine papilloma virus, avian sarcoma virus, cytomegalovirus (CMV), a retrovirus and Simian Virus 40 (SV40), from heterologous mammalian promoters such as the actin promoter or a very strong promoter, e.g. a ribosomal protein promoter, and from the promoter normally associated with RNA 3'-terminal phosphate cyclase sequence, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding RNA 3'-terminal phosphate cyclase by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are orientation and position independent. Many enhancer sequences are known from mammalian genes (e.g. elastase and globin). However, typically one will employ an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270) and the CMV early promoter enhancer. The enhancer may be spliced into the vector at a position 5' or 3' to RNA 3'-terminal phosphate cyclase DNA, but is preferably located at a site 5' from the promoter.

The polypeptide according to the invention may advantageously be expressed in insect cell systems. Insect cells suitable for use in the method of the invention include, in principle, any lepidopteran cell which is capable of being transformed with an expression vector and expressing heterologous proteins encoded thereby. In particular, use of the Sf cell lines, such as the *Spodoptera frugiperda* cell line IPBL-SF-21 AE (Vaughn *et al.*, (1977) In Vitro, 13, 213-217) is preferred. The derivative cell line Sf9 is particularly preferred. However, other cell lines, such as *Tricoplusia ni* 368 (Kurstack and Marmorosch, (1976) Invertebrate Tissue Culture Applications in Medicine, Biology and Agriculture. Academic Press, New York, USA) may be employed. These cell lines, as well as other insect cell lines

suitable for use in the invention, are commercially available (e.g. from Stratagene, La Jolla, CA, USA).

As well as expression in insect cells in culture, the invention also comprises the expression of heterologous proteins in whole insect organisms. The use of virus vectors such as baculovirus allows infection of entire insects, which are in some ways easier to grow than cultured cells as they have fewer requirements for special growth conditions. Large insects, such as silk moths, provide a high yield of heterologous protein. The protein can be extracted from the insects according to conventional extraction techniques.

Expression vectors suitable for use in insect expression systems according to the invention include all vectors which are capable of expressing foreign proteins in insect cell lines. In general, vectors which are useful in mammalian and other eukaryotic cells are also applicable to insect cell culture. Baculovirus vectors, specifically intended for insect cell culture, are especially preferred and are widely obtainable commercially (e.g. from Invitrogen and Clontech). Other virus vectors capable of infecting insect cells are known, such as Sindbis virus (Hahn *et al.*, (1992) PNAS (USA) 89, 2679-2683). The baculovirus vector of choice (reviewed by Miller (1988) Ann. Rev. Microbiol. 42, 177-199) is *Autographa californica* multiple nuclear polyhedrosis virus, AcMNPV.

Typically, the heterologous gene replaces at least in part the polyhedrin gene of AcMNPV, since polyhedrin is not required for virus production. In order to insert the heterologous gene, a transfer vector is advantageously used. Transfer vectors are prepared in *E. coli* hosts and the DNA insert is then transferred to AcMNPV by a process of homologous recombination.

Advantageously, a eukaryotic expression vector encoding RNA 3'-terminal phosphate cyclase may comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell chromatin, which is of importance especially where the RNA 3'-terminal phosphate cyclase gene is to be expressed in the context of a permanently-transfected eukaryotic cell line in which chromosomal integration of the vector has occurred, in vectors designed for gene therapy applications or in transgenic animals.

The invention also provides a transgenic non-human mammal which has been modified to modulate the expression of endogenous RNA 3'-terminal phosphate cyclase. Preferably, the transgenic non-human mammal is a transgenic mouse. For example, therefore, a transgenic mouse may be designed in which RNA 3'-terminal phosphate cyclase production is greatly reduced or eliminated Alternatively, the transgenic mouse of

the invention may express elevated levels of RNA 3'-terminal phosphate cyclase or may be subject to regulation of RNA 3'-terminal phosphate cyclase expression in a developmentally or tissue-specific manner, or via control by exogenous agents. Study of such an animal provides insights into the importance of RNA 3'-terminal phosphate cyclase *in vivo*.

In accordance with yet another embodiment of the present invention, there are provided antibodies specifically recognising and binding to RNA 3'-terminal phosphate cyclase. For example, such antibodies may be generated against the RNA 3'-terminal phosphate cyclase having the amino acid sequences set forth in SEQ ID No. 2. Alternatively, RNA 3'-terminal phosphate cyclase or RNA 3'-terminal phosphate cyclase fragments (which may also be synthesised by *in vitro* methods) are fused (by recombinant expression or an *in vitro* peptidyl bond) to an immunogenic polypeptide and this fusion polypeptide, in turn, is used to raise antibodies against an RNA 3'-terminal phosphate cyclase epitope.

Anti-RNA 3'-terminal phosphate cyclase antibodies are recovered from the serum of immunised animals. Alternatively, monoclonal antibodies are prepared from cells *in vitro* or from *in vivo* immunised animals in conventional manner.

The antibodies of the invention are useful for studying RNA 3'-terminal phosphate cyclase tissue localisation, screening of an expression library to identify nucleic acids encoding RNA 3'-terminal phosphate cyclase or the structure of functional domains, as well as for the purification of RNA 3'-terminal phosphate cyclase, and the like.

Antibodies according to the invention may be whole antibodies of natural classes, such as IgE and IgM antibodies, but are preferably IgG antibodies. Moreover, the invention includes antibody fragments, such as Fab, $F(ab')_2$, Fv and ScFv. Small fragments, such Fv and ScFv, possess advantageous properties for diagnostic and therapeutic applications on account of their small size and consequent superior tissue distribution.

The antibodies according to the invention are especially indicated for diagnostic and therapeutic applications. Accordingly, they may be altered antibodies comprising an effector protein such as a toxin or a label. Especially preferred are labels which allow the imaging of the distribution of the antibody *in vivo*. Such labels may be radioactive labels or radioopaque labels, such as metal particles, which are readily visualisable within the body of a patient. Moreover, the may be fluorescent labels or other labels which are visualisable on tissue samples removed from patients.

Recombinant DNA technology may be used to improve the antibodies of the invention. Thus, chimeric antibodies may be constructed in order to decrease the

immunogenicity thereof in diagnostic or therapeutic applications. Moreover, immunogenicity may be minimised by humanising the antibodies by CDR grafting [see European Patent Application 0 239 400 (Winter)] and, optionally, framework modification [see international patent application WO 90/07861 (Protein Design Labs)].

Antibodies according to the invention may be obtained from animal serum, or, in the case of monoclonal antibodies or fragments thereof, produced in cell culture. Recombinant DNA technology may be used to produce the antibodies according to established procedure, in bacterial or preferable in mammalian cell culture. The selected cell culture system preferably secretes the antibody product.

Therefore, the present invention includes a process for the production of an antibody according to the invention comprising culturing a host, e.g. *E. coli* or a mammalian cell, which has been transformed with a hybrid vector comprising an expression cassette comprising a promoter operably linked to a first DNA sequence encoding a signal peptide linked in the proper reading frame to a second DNA sequence encoding said protein, and isolating said protein.

Multiplication of hybridoma cells or mammalian host cells *in vitro* is carried out in suitable culture media, which are the customary standard culture media, for example Dulbecco's Modified Eagle Medium (DMEM) or RPMI 1640 medium, optionally replenished by a mammalian serum, e.g. fetal calf serum, or trace elements and growth sustaining supplements, e.g. feeder cells such as normal mouse peritoneal exudate cells, spleen cells, bone marrow macrophages, 2-aminoethanol, insulin, transferrin, low density lipoprotein, oleic acid, or the like. Multiplication of host cells which are bacterial cells or yeast cells is likewise carried out in suitable culture media known in the art, for example for bacteria in medium LB, NZCYM, NZYM, NZM, Terrific Broth, SOB, SOC, 2 x YT, or M9 Minimal Medium, and for yeast in medium YPD, YEPD, Minimal Medium, or Complete Minimal Dropout Medium.

*In vitro* production provides relatively pure antibody preparations and allows scale-up to give large amounts of the desired antibodies. Techniques for bacterial cell, yeast or mammalian cell cultivation are known in the art and include homogeneous suspension culture, e.g. in an airlift reactor or in a continuous stirrer reactor, or immobilised or entrapped cell culture, e.g. in hollow fibres, microcapsules, on agarose microbeads or ceramic cartridges.

Large quantities of the desired antibodies can also be obtained by multiplying mammalian cells *in vivo*. For this purpose, hybridoma cells producing the desired

antibodies are injected into histocompatible mammals to cause growth of antibody-producing tumours. Optionally, the animals are primed with a hydrocarbon, especially mineral oils such as pristane (tetramethyl-pentadecane), prior to the injection. After one to three weeks, the antibodies are isolated from the body fluids of those mammals. For example, hybridoma cells obtained by fusion of suitable myeloma cells with antibody-producing spleen cells from Balb/c mice, or transfected cells derived from hybridoma cell line Sp2/0 that produce the desired antibodies are injected intraperitoneally into Balb/c mice optionally pre-treated with pristane, and, after one to two weeks, ascitic fluid is taken from the animals.

The cell culture supernatants are screened for the desired antibodies, preferentially by immunofluorescent staining of cells expressing RNA 3'-terminal phosphate cyclase, by immunoblotting, by an enzyme immunoassay, e.g. a sandwich assay or a dot-assay, or a radioimmunoassay.

For isolation of the antibodies, the immunoglobulins in the culture supernatants or in the ascitic fluid may be concentrated, e.g. by precipitation with ammonium sulphate, dialysis against hygroscopic material such as polyethylene glycol, filtration through selective membranes, or the like. If necessary and/or desired, the antibodies are purified by the customary chromatography methods, for example gel filtration, ion-exchange chromatography, chromatography over DEAE-cellulose and/or (immuno-)affinity chromatography, e.g. affinity chromatography with RNA 3'-terminal phosphate cyclase protein or with Protein-A.

The invention further concerns hybridoma cells secreting the monoclonal antibodies of the invention. The preferred hybridoma cells of the invention are genetically stable, secrete monoclonal antibodies of the invention of the desired specificity and can be activated from deep-frozen cultures by thawing and recloning.

The invention also concerns a process for the preparation of a hybridoma cell line secreting monoclonal antibodies directed RNA 3'-terminal phosphate cyclase, characterised in that a suitable mammal, for example a Balb/c mouse, is immunised with purified RNA 3'-terminal phosphate cyclase protein, an antigenic carrier containing purified RNA 3'-terminal phosphate cyclase or with cells bearing RNA 3'-terminal phosphate cyclase, antibody-producing cells of the immunised mammal are fused with cells of a suitable myeloma cell line, the hybrid cells obtained in the fusion are cloned, and cell clones secreting the desired antibodies are selected. For example spleen cells of Balb/c mice immunised with cells bearing RNA 3'-terminal phosphate cyclase are fused with cells of the myeloma cell line PAI

or the myeloma cell line Sp2/0-Ag14, the obtained hybrid cells are screened for secretion of the desired antibodies, and positive hybridoma cells are cloned.

Preferred is a process for the preparation of a hybridoma cell line, characterised in that Balb/c mice are immunised by injecting subcutaneously and/or intraperitoneally between 10 and $10^7$ and $10^8$ cells of human tumour origin which express RNA 3'-terminal phosphate cyclase containing a suitable adjuvant several times, e.g. four to six times, over several months, e.g. between two and four months, and spleen cells from the immunised mice are taken two to four days after the last injection and fused with cells of the myeloma cell line PAI in the presence of a fusion promoter, preferably polyethylene glycol. Preferably the myeloma cells are fused with a three- to twentyfold excess of spleen cells from the immunised mice in a solution containing about 30 % to about 50 % polyethylene glycol of a molecular weight around 4000. After the fusion the cells are expanded in suitable culture media as described hereinbefore, supplemented with a selection medium, for example HAT medium, at regular intervals in order to prevent normal myeloma cells from overgrowing the desired hybridoma cells.

The invention also concerns recombinant DNAs comprising an insert coding for a heavy chain variable domain and/or for a light chain variable domain of antibodies directed to RNA 3'-terminal phosphate cyclase as described hereinbefore. By definition such DNAs comprise coding single stranded DNAs, double stranded DNAs consisting of said coding DNAs and of complementary DNAs thereto, or these complementary (single stranded) DNAs themselves.

Furthermore, DNA encoding a heavy chain variable domain and/or for a light chain variable domain of antibodies directed RNA 3'-terminal phosphate cyclase can be enzymatically or chemically synthesised DNA having the authentic DNA sequence coding for a heavy chain variable domain and/or for the light chain variable domain, or a mutant thereof. A mutant of the authentic DNA is a DNA encoding a heavy chain variable domain and/or a light chain variable domain of the above-mentioned antibodies in which one or more amino acids are deleted or exchanged with one or more other amino acids. Preferably said modification(s) are outside the CDRs of the heavy chain variable domain and/or of the light chain variable domain of the antibody. Such a mutant DNA is also intended to be a silent mutant wherein one or more nucleotides are replaced by other nucleotides with the new codons coding for the same amino acid(s). Such a mutant sequence is also a degenerated sequence. Degenerated sequences are degenerated within the meaning of the genetic code in that an unlimited number of nucleotides are replaced by other

nucleotides without resulting in a change of the amino acid sequence originally encoded. Such degenerated sequences may be useful due to their different restriction sites and/or frequency of particular codons which are preferred by the specific host, particularly E. coli, to obtain an optimal expression of the heavy chain murine variable domain and/or a light chain murine variable domain.

The term mutant is intended to include a DNA mutant obtained by in vitro mutagenesis of the authentic DNA according to methods known in the art.

For the assembly of complete tetrameric immunoglobulin molecules and the expression of chimeric antibodies, the recombinant DNA inserts coding for heavy and light chain variable domains are fused with the corresponding DNAs coding for heavy and light chain constant domains, then transferred into appropriate host cells, for example after incorporation into hybrid vectors.

The invention therefore also concerns recombinant DNAs comprising an insert coding for a heavy chain murine variable domain of an antibody directed RNA 3'-terminal phosphate cyclase fused to a human constant domain $\gamma$, for example $\gamma$1, $\gamma$2, $\gamma$3 or $\gamma$4, preferably $\gamma$1 or $\gamma$4. Likewise the invention concerns recombinant DNAs comprising an insert coding for a light chain murine variable domain of an antibody directed to RNA 3'-terminal phosphate cyclase fused to a human constant domain $\kappa$ or $\lambda$, preferably $\kappa$.

In another embodiment the invention pertains to recombinant DNAs coding for a recombinant DNA wherein the heavy chain variable domain and the light chain variable domain are linked by way of a DNA insert coding for a spacer group, optionally comprising a signal sequence facilitating the processing of the antibody in the host cell and/or a DNA coding for a peptide facilitating the purification of the antibody and/or a DNA coding for a cleavage site and/or a DNA coding for a peptide spacer and/or a DNA coding for an effector molecule.

The DNA coding for an effector molecule is intended to be a DNA coding for the effector molecules useful in diagnostic or therapeutic applications. Thus, effector molecules which are toxins or enzymes, especially enzymes capable of catalysing the activation of prodrugs, are particularly indicated. The DNA encoding such an effector molecule has the sequence of a naturally occurring enzyme or toxin encoding DNA, or a mutant thereof, and can be prepared by methods well known in the art.

Antibodies and antibody fragments according to the invention are useful in diagnosis and therapy. Accordingly, the invention provides a composition for therapy or diagnosis comprising an antibody according to the invention.

In the case of a diagnostic composition, the antibody is preferably provided together with means for detecting the antibody, which may be enzymatic, fluorescent, radioisotopic or other means. The antibody and the detection means may be provided for simultaneous, simultaneous separate or sequential use, in a diagnostic kit intended for diagnosis.

The invention is further described, for the purpose of illustration only, in the following examples:

**Example 1**

**Cloning of human RNA 3'-terminal phosphate cyclase**

Protein purification and peptide sequencing: The cyclase is purified from HeLa cells following an established protocol (Vicente, O. and Filipowicz, W. (1988) Eur. J. Biochem. 176, 431-439) with some modifications. (1) The cyclase is eluted from the Heparin-Sepharose column (step 3) with a linear gradient of 75-450 mM NaCl. Fractions corresponding to pool (a) (Filipowicz *et al.* (1985) PNAS (USA) 82, 1316-1320) eluting at 260 mM NaCl are collected and applied, after dialysis, to the polyA-Sepharose column (Vicente and Filipowicz, 1988). (2) The mono-S step (step 5) is omitted. The polyA-Sepharose fraction, concentrated and dialysed against buffer C containing 75 mM NaCl, is directly applied to the Blue-Sepharose column. Material eluting at 0.6 M NaCl (fraction BS-600; Vicente and Filipowicz, 1988) is succesively concentrated in Centricon-30 and Microcon-10 (Amicon). 10 µg of the protein is applied to SDS-PAGE 10% gel. Proteins are blotted into nitrocellulose (Scleicher & Schuell) and stained with Ponceau S. The 39 kDa band, representing the cyclase, is excised and treated with trypsin. Proteolytic peptides are resolved by HPLC and sequenced by Dr. W.S. Lane (Harvard MicroChem, Cambridge, MA). Sequences of four peptides are established (*pep1* VEVDGSIMEGGGQIL, *pep2* GYYPK, *pep3* QLNPINLTER, *pep4* DLYVNIQPVQE).

Cloning of the cyclase cDNA: Partial sequence of the cyclase cDNA is obtained by two consecutive PCR amplifications, using λgt11 human HeLa cell cDNA library (Clontech) as a template. For the first PCR reaction oligonucleotides **1** (5'-CACATGGCTGAATATCGAC-3') SEQ ID No. 12, corresponding to the border sequence of

the *Eco*RI λgt11 cloning site and **2** (5'-TC(C/T)TG(C/G)ACA/G/T)GG(C/T)TG(A/G)AT (A/G)TT-3'), SEQ ID No. 13, representing a mixture of 96 oligomers complementary to the sequence encoding sequence NIQPVQE of peptide *pep4* are used as upstream and downstream primers, respectively. 100 µl PCR reaction contained 1 µM oligonucleotide **1**, 4 µM oligonucleotide **2**, 200 ng of the library DNA, 250 µM of each of the four nucleotides, 2.5 U Taq DNA polymerase in the PERKIN ELMER buffer. Thirty PCR cycles consisted of denaturation at 94°C for 40 sec, annealing at 45°C for 1 min, and extension at 70°C for 1.5 min each cycle. For the second PCR reaction, oligonucleotides **3** (5'-GA(C/T)GG(A/C/G/T)TC(A/C/T)AT(A/C/T)ATGGA (A/G)GG-3'), SEQ ID No. 14, a mixture of 144 oligomers coding for the sequence DGSIMEG of peptide *pep1* and **4** (5'-GGCTG(A/G)AT(A/G)TT(A/C/G)AC(A/G)TA(C/G)AG (A/G)TC-3'), SEQ ID No. 15, overlapping with oligonucleotide **2** is a mixture of 96 oligomers coding for the sequence DLYVNIQP of peptide *pep4*) are used as primers. The 100 µl PCR reaction contained 4 µM of each primer, 250 µM of each of the four nucleotides, 2.5 U Taq DNA polymerase and 8 µl of the first PCR reaction. Thirty PCR cycles consisted of denaturation at 94°C for 40 sec, annealing at 52°C for 1 min, and extension at 72°C for 1.5 min each cycle. A 668 bp amplified DNA fragment is subcloned into the *Sma*I site of pBluescribe vector (Stratagene) and sequenced by the dideoxy chain termination method of Sanger *et al.* (Proc. Natl. Acad. Sci. USA 74, 5463-5467,1977). The fragment contained a large ORF encoding the peptides *pep2* and *pep3*. The PCR amplification scheme described above represents one of many similar reactions tested with different combination of primers; it is the only one which yielded the cDNA encoding expected peptide sequences.

In order to get a full lenth cDNA, the HeLa λgt11 cDNA library (Clontech) is screened with the PCR amplified fragment as a probe. Hybridizations are performed overnight at 42°C in 5 x SSPE, 50% formamide, 5 x Denhart's solution, 1% SDS and 50 µg/ml denatured salmon sperm DNA. Eight clones are isolated after screening one million recombinant phages. All clones are subcloned in pBluescribe vector (Stratagene) and analysed by restriction mapping and sequencing of the ends. The longest clones are sequenced on both strands.

**Northern blot analysis**

Northern blot analysis is done using formaldehyde-agarose gels and 15 µg of total RNA from various cell lines (Ausubel *et al.*, (1990) Current Protoclos in Molecular Biology

(New York: Greene Publishing). The RNA is blotted onto GeneScreen membranes and UV cross-linked. The blot containing RNAs originating from different human organs is purchased from Clontech. The integrity as well as the amount of RNAs is checked by using a human β-actin cDNA, furnished by Clontech, as control probe. The probes are labelled with [$\alpha$-$^{32}$P]dCTP (300 Ci/mmol, Amersham) by the random priming method (Feinberg and Vogelstein, 1983). The hybridization is for 16 h at 42°C in 5 x SSPE, 50% formamide, 10% dextran sulfate, 1% SDS and 50 μg/ml denatured salmon sperm DNA. The blots are subsequently washed in 2 x SSC and 0.1% SDS for 30 min at 42°C, and 0.2 x SSC and 0.1% SDS for 30 min at 42°C and then at 60°C.

**Overexpression of the cyclase in *E. coli***

The *Bam*HI sites are introduced 5' and 3' of the cyclase coding sequence using the PCR site-directed mutagenesis. The *Bam*HI/*Bam*HI fragment is cloned into pGEX-2T vector (Pharmacia) for expression in the *E. coli* strain BL21(DE3). In this construct, the *Schistosoma japonicum* glutathione S-transferase is placed in frame at the N-terminus of the fusion protein. The protein remained soluble during expression in *E. coli* and is purified in the native form under non-denaturing conditions on the Glutathione Sepharose 4B resin (following the manufacturer's protocol; Pharmacia). No detectable contaminating proteins are apparent after Coomassie blue staining of the protein gel. The purified cyclase is then applied to a 10 ml Sephadex G-25 column equilibrated and eluted with 30 mM HEPES-KOH (pH 7.6), 0.1 mM EDTA, 0.5 mM DTT, 5% (v/v) glycerol, 0.01% Triton X-100 and 10 μM PMSF. Protein concentration is measured by the method of Bradford (1976) using the reagent obtained from BioRad and BSA as a standard.

**Assays of cyclase and chromatography**

Preparation of substrates: Two synthetic oligoribonucleotides (oligoSY: 5'-CCCCACCCCG-3'), SEQ ID No. 16, and (oligoSR: 5'-AAAATAAAAG-3'), SEQ ID No. 17, are obtained from MWG-BIOTECH (Munich). They are 3'-terminally labelled using [5'-$^{32}$P]pCp (*pCp) and T4 RNA ligase, to produce (Np)$_9$Gp*Cp. Reaction mixtures (15 μl) contained 70 mM Hepes/KOH, pH 8.3, 10 mM $MgCl_2$, 3 mM dithiothreitol, 10% glycerol, 10% dimethylsulfoxide, 40 μM ATP, 1.1 μM p*Cp (spec. 3000 Ci/mmol), 0.7 μg of oligoribonucleotides and 6 units of T4 RNA ligase. After 14 h at 4°C, samples are diluted with 60 μl of 50 mM Hepes/KOH, pH 7.6, containing 3 mM EDTA and incubated at 37°C for

1 h with 5 U of RNase T1. SDS is then added to a final concentration of 0.1% and the samples extracted with phenol/chloroform/isoamyl alcohol (25:24:1). The substrate in the aqueous phase, corresponding to $(Np)_9Gp^*$, is purified on Sephadex G-25 (fine) spin column pre-equilibrated with 20 mM ammonium acetate, pH 5.5. Radioactive substrates are analysed by digestion with RNase T2, nuclease P1 or Calf Intestinal Phosphatase (CIP), followed by TLC on cellulose plates in solvent A (saturated $(NH_4)_2SO_4$, 3 M Na acetate, isopropylalcohol (80:6:2)). Over 90% of the label is always present as the 3'-terminal $G^{3'}p^*$.

Cyclase assay: The cyclase activity is assayed by the Norit method as described before (Filipowicz and Vicente; 1990). Analysis of the reaction products by TLC is as described above.

**Labelling of the RNA 3'-terminal phosphate cyclase with [$\alpha$-$^{32}$P]ATP**

Overexpressed and purified human cyclase is labelled with [$\alpha$-$^{32}$P]ATP. 15 µl of the reaction containing 50 ng cyclase, 200 µM [$\alpha$-$^{32}$P]ATP (1500 Ci/mmol), are incubated under standard cyclase conditions, except that the $(Np)_nGp^*$ substrate is omitted. After 30 min at 25°C, the products of the reaction are analysed by SDS/PAGE and autoradiography. Immediatly before being applied to the gel, samples are supplemented with unlabelled ATP (final concentration 10 mM).

**Transfection of HeLa cells and analysis by indirect immunofluorescence**

Construction of pBact-CYC-myc: The *Eco*RI sites are introduced 5' and 3' of the cyclase coding sequence by the PCR based site directed mutagenesis using oligonucleotides **5** (5'-GAGGAGAAAGAATTCATGGCGGGG CCGTGG-3'), SEQ ID No. 18, and **6** (5'-AGTGGTGATGGTGGAATTCCTATAGATTTG-3'), SEQ ID No. 19, as upstream and downstream primers, respectively. The *Eco*RI/*Eco*RI fragment is cloned into the pBact-myc vector (Cravchik and Matus, 1993) for expression of the myc-epitope-tagged cyclase in transfected HeLa cells.

Transfection of HeLa cells and immunolocalisation: Hela cells are cultured in the DME medium (Sambrook et al., Molecular Cloning. A Laboratory Manual (2nd ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) with the addition of 10% fetal calf serum in 5% $CO_2$ humidified atmosphere. Cell are transfected with pBact-CYC-myc (2-10 µg), using the Calcium phosphate transformation method (Chen and Okayama, 1987) and fixed with 3% paraformaldehyde in PBS (Sambrook *et al.*, 1989) at room temperature for 20

min. Cells are permeabilised for 10 min with 0.5% Triton X-100 solution in PBS, and blocked with 5% non immune goat serum in PBS for 20 min. The myc-tagged cyclase is immunoprobed with a mAb GE10 raised to the human c-myc epitope (Evan *et al.*, (1985) Mol. Cell. Biol. 5, 3610-3616), obtained from the European Collection of Cell Cultures (Porton Down, UK). The FITC-conjugated goat anti-mouse antibody (AffiniPure F(ab')2 fragment, Jackson ImmunoResearch Laboratories, Inc.) is used as a secondary antibody. Samples are examined with the Zeiss Axiophot microscop and the Leica confocal scanning laser microscope using a 63X objective. Images are recorded using Leica SCANware 4.2 provided with the system.

**Example 2**

**Cloning of *E. coli* RNA 3'-terminal phosphate cyclase**

Two neighbouring ORFs of unknown function, present on the *E. coli* K-12 chromosome (the GenEMBL accession number U18997) are identified by the data base searching using the human cyclase protein sequence. The region of 1047 bp (positions 336558 to 337605 in U18997), coverring the two ORFs, is PCR amplified using 200 ng of the *E coli* genomic DNA as template and prepared as described (Ausubel *et al.*, 1990). Re-sequencing of this region showed several sequencing errors which result in change of the reading frames. This chromosomal region encodes a single ORF showing similarity to the human RNA 3'-terminal phosphate cyclase.

**Overexpression of the cyclase in *E. coli***

*Nco*I and *Bam*HI sites are introduced 5' and 3', respectively of the cyclase coding sequence by the PCR based site directed mutagenesis using oligonucleotides **1** (5'-TGCGTAAAAGGATCCATGGTGAAAAGGATG-3'), SEQ ID No. 20, and **8** (5'-CTTACATCTCTGGATCCTTCAATGCTCAC-3'), SEQ ID No. 21, as upstream and downstream primers, respectively. The *Nco*I/*Bam*HI fragment is cloned into the pET-11d vector (Novagen) modified in order to include 6 histidine residues in frame at the C-terminus of the recombinant protein. The protein overexpression is performed in the *E. coli* strain BL21(DE3). The protein remained soluble and could be purified in the native form under non-denaturing conditions using the Ni-NTA resin (according to the protocol provided by

QIAGEN). No detectable contaminating proteins are visible after the Coomassie blue staining of the protein gels. The purified cyclase is applied to a 10 ml Sephadex G-25 column equilibrated and eluted with 30 mM HEPES-KOH (pH 7.6), 0.1 mM EDTA, 0.5 mM DTT, 5% (v/v) glycerol, 0.01% Triton X-100 and 10 µM PMSF. Protein concentration is measured by the method of Bradford, M.M. (1976) Anal. Biochem. 72, 248-254 using the reagent obtained from BioRad and BSA as a standard.

**The cyclase knockout in *E. coli***

Bacterial strain and plasmid: The gene replacement is performed in the strain MC1061 [F⁻ *ara*D139 (*ara-leu*)7697 (*lac*)X74 *gal*U *gal*K *str*A]. pMAK705 is a chloramphenicol-resistant derivative of pSC101 that contains a temperature-sensitive rep mutation which prevents from the replication at 44°C (Hamilton *et al.*, 1989).

Construction of pMAK705*cyc*⁻::*kan*$^r$ plasmid: the 1324 bp *Sma*I/*Hin*dIII fragment from pREP4 (Qiagen), which includes the kanamycin resistance gene is ligated into *Sma*I/*Hin*dIII sites of pBluescriptKS (Stratagene), yielding pBSKan$^r$. A *Xba*I/*Bam*HI fragment of 669 bp corresponding to the cyclase gene 3' flanking sequences is obtained using the PCR site-directed mutagenesis with oligonucleotides **9** (5'-ATCTCTAGAGATACCTGTTGCTGCTTA-ATC-3'), SEQ ID No. 22, and **10** (5'-GGTCGCGGATCCCTCATGCCATCTGCTGAC-3'), SEQ ID No. 23, as upstream and downstream primers, respectively and 200 ng of *E coli* genomic DNA. The *Xba*I/*Bam*HI fragment is cloned into *Xba*I/*Bam*HI sites of pBSKan$^r$ polycloning region; the resulting plasmid is called pBS3'(669)/Kan$^r$. A *Cla*I/*Sal*I frangment of 620 bp corresponding to the cyclase gene 5' flanking sequences is obtained using the PCR site-directed mutagenesis with oligonucleotides **11** (5'-TGCGCCATCGATCGCAATCATC CTTTTCATC-3'), SEQ ID No. 24, and **12** (5'-CTTACTTTGTCGACCTGGCACAAAAAG AGATG-3'), SEQ ID No. 25, as upstream and downstream primers, respectively and 200 ng of the *E coli* genomic DNA. The *Cla*I/*Sal*I fragment is cloned into *Cla*I/*Sal*I sites of the pBS3'(669)/Kan$^r$ polycloning region, yielding pBS3'(669)/Kan$^r$/5'(620). Finally, the 2.6 kb *Xba*I/*Sal*I fragment from pBSKan$^r$/3'(669)/5'(620) is ligated into XbaI/SalI sites of pMAK705, yielding pMAK705*cyc*⁻::*kan*$^r$.

Cyclase chromosomal gene replacement: Chromosomal insertion and excision steps with pMAK705*cyc*⁻::*kan*$^r$, using chloramphenicol as a selection marker, are performed in MC1061 as described by Hamilton *et al.* (1989), with the following modifications. MC1061 is transformed and plated at 30°C. Several transformants are grown up at 30°C in a rich

medium (2YT) in the presence of 30 μg/ml chloramphenicol to $10^8$ cells/ml. Different dilutions are plated at both 30°C (for viable count) and 44°C (for cointegrates). The observed insertion frequency of $3 \times 10^{-4}$ is consistent with the results of Hamilton *et al.* (1989) obtained with homologous sequences of comparable size. For excision of the plasmid, 9 different colonies are grown individually overnight at 30°C for three days, re-inoculating daily into the fresh medium containing chloramphenicol. Since the plasmid pMAK705*cyc*$^-$::*kan*$^r$ is not designed for complementation in *trans* after resolution, the gene replacement could have only be observed if the cyclase gene are not essential. The cultures are streaked out on LB plates containing 25 μg/ml kanamycine at 44°C. Three individual colonies out of 45 analysed, generated chloramphenicol-sensitive clones at 44°C, indicating that the excision of the plasmid had been successful. Subsequent experiments are work done with only one of the strains. The resident plasmid is cured by growing the cells at 44°C on LB plates. Different PCR reactions, using MC1061 as control strain, confirmed that the gene replacement had taken place.

**Example 3**

**Phylogenic relationship of the human cyclase with other ORFs.**

The primary sequence of the human cyclase protein is used for database "homology" searching. Several ORFs of unknown functions from different species, gave high scores over 100, using TBLASTN Program. The highest similarities are found in two adjacent ORFs from *E. coli* K12. The chromosomal region is re-sequenced and showed several sequencing errors. Based on our sequencing only one large ORF is present in this region, which is 55% similar and 32% identical to the human cyclase. The bacterial protein is overexpressed in *E. coli* and the RNA 3'-terminal phosphate cyclase activity is detected.

Other ORFs, localised in ESTs or chromosomal regions, with significant similarities to the human cyclase emerged from different sequencing programs:

| | |
|---|---|
| Ø human: | Several different ESTs encoding a 2nd cyclase-like protein |
| Ø *Danio rerio* (Zebra fish) | One EST |

| | |
|---|---|
| Ø Drosophila | Chromosomal sequence |
| Ø Arabidopsis | one EST |
| Ø *S. cerevisiae* | Chromosomal sequence |
| Ø *S. pombe* | Chromosomal sequence |

The phylogenic relationship of all cyclase and cyclase-like proteins is determined on a cumputer generated dendrogram, using the PileUp GCG program.

All sequences clearly fit into two (or perhaps three) subfamilies and they are probably derived from a common ancestor.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: NOVARTIS AG
(B) STREET: SCHWARZWALDALLEE
(C) CITY: Basel
(E) COUNTRY: Switzerland
(F) POSTAL CODE (ZIP): 4002
(G) TELEPHONE: +41 61 69 11 11
(H) TELEFAX: + 41 61 696 79 76
(I) TELEX: 962 991

(ii) TITLE OF INVENTION: Protein

(iii) NUMBER OF SEQUENCES: 25

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1530 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(C) INDIVIDUAL ISOLATE: HeLa

(vii) IMMEDIATE SOURCE:
(B) CLONE: RNA cyclase 1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:152..1252

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:152..1252

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAGAGGCGCC GCTTCTTCCG CTTTCTCGTC AGGCTCCTGC AACCCCAGGC ATGAACCAAG 60

GTTTCTGAAC TACTGGGCGG GACCAACGTC TCTTCTTTCT CCGCTCTGGC GGAGGCTTTG 120

TCGCTGCGGG CTGGGCCCCA GGGTGTCCCC C ATG GCG GGG CCG TGG GTG GAG 172
Met Ala Gly Pro Trp Val Glu
1 5

GTC GAT GGC AGC ATC ATG GAA GGG GGC GGC CAG ATC CTG AGA GTC TCT 220
Val Asp Gly Ser Ile Met Glu Gly Gly Gly Gln Ile Leu Arg Val Ser
10 15 20

ACG GCC TTG AGC TGT CTC CTA GGC CTC CCC TTG CGG GTG CAG AAG ATC 268
Thr Ala Leu Ser Cys Leu Leu Gly Leu Pro Leu Arg Val Gln Lys Ile
25 30 35

CGA GCC GGC CGG AGC ACG CCA GGC CTG AGG CCT CAA CAT TTA TCT GGA 316
Arg Ala Gly Arg Ser Thr Pro Gly Leu Arg Pro Gln His Leu Ser Gly
40 45 50 55

CTG GAA ATG ATT CGA GAT TTG TGT GAT GGG CAA CTG GAG GGG GCA GAA 364
Leu Glu Met Ile Arg Asp Leu Cys Asp Gly Gln Leu Glu Gly Ala Glu
60 65 70

ATT GGC TCA ACA GAA ATA ACC TTT ACA CCA GAG AAG ATC AAA GGT GGA 412
Ile Gly Ser Thr Glu Ile Thr Phe Thr Pro Glu Lys Ile Lys Gly Gly
75 80 85

```
ATC CAC ACA GCA GAT ACC AAG ACA GCA GGG AGT GTG TGC CTC TTG ATG          460
Ile His Thr Ala Asp Thr Lys Thr Ala Gly Ser Val Cys Leu Leu Met
         90                  95                 100

CAG GTC TCA ATG CCG TGT GTT CTC TTT GCT GCT TCT CCA TCA GAA CTT          508
Gln Val Ser Met Pro Cys Val Leu Phe Ala Ala Ser Pro Ser Glu Leu
    105                 110                 115

CAT TTG AAA GGT GGA ACT AAT GCT GAA ATG GCA CCA CAG ATC GAT TAT          556
His Leu Lys Gly Gly Thr Asn Ala Glu Met Ala Pro Gln Ile Asp Tyr
120                 125                 130                 135

ACA GTG ATG GTC TTC AAG CCA ATT GTT GAA AAA TTT GGT TTC ATA TTT          604
Thr Val Met Val Phe Lys Pro Ile Val Glu Lys Phe Gly Phe Ile Phe
                140                 145                 150

AAT TGT GAC ATT AAA ACA AGG GGA TAT TAC CCA AAA GGG GGT GGT GAA          652
Asn Cys Asp Ile Lys Thr Arg Gly Tyr Tyr Pro Lys Gly Gly Gly Glu
           155                 160                 165

GTG ATT GTT CGA ATG TCA CCA GTT AAA CAA TTG AAC CCT ATA AAT TTA          700
Val Ile Val Arg Met Ser Pro Val Lys Gln Leu Asn Pro Ile Asn Leu
         170                 175                 180

ACT GAG CGT GGC TGT GTG ACT AAG ATA TAT GGA AGA GCT TTC GTT GCT          748
Thr Glu Arg Gly Cys Val Thr Lys Ile Tyr Gly Arg Ala Phe Val Ala
    185                 190                 195

GGT GTT TTG CCA TTT AAA GTA GCA AAA GAT ATG GCA GCG GCA GCA GTT          796
Gly Val Leu Pro Phe Lys Val Ala Lys Asp Met Ala Ala Ala Ala Val
200                 205                 210                 215

AGA TGC ATC AGA AAG GAG ATC CGG GAT TTG TAT GTT AAC ATC CAG CCT          844
Arg Cys Ile Arg Lys Glu Ile Arg Asp Leu Tyr Val Asn Ile Gln Pro
                220                 225                 230

GTT CAA GAA CCT AAA GAC CAA GCA TTT GGC AAT GGA AAT GGA ATA ATA          892
Val Gln Glu Pro Lys Asp Gln Ala Phe Gly Asn Gly Asn Gly Ile Ile
           235                 240                 245

ATT ATT GCT GAG ACC TCC ACT GGC TGT TTG TTT GCT GGA TCA TCG CTT          940
Ile Ile Ala Glu Thr Ser Thr Gly Cys Leu Phe Ala Gly Ser Ser Leu
         250                 255                 260
```

```
GGT AAA CGA GGT GTA AAT GCA GAC AAA GTT GGA ATT GAA GCT GCC GAA      988
Gly Lys Arg Gly Val Asn Ala Asp Lys Val Gly Ile Glu Ala Ala Glu
    265                 270                 275

ATG CTA TTA GCA AAT CTT AGA CAT GGT GGT ACT GTG GAT GAG TAT CTG     1036
Met Leu Leu Ala Asn Leu Arg His Gly Gly Thr Val Asp Glu Tyr Leu
280                 285                 290                 295

CAA GAC CAG CTG ATT GTT TTC ATG GCA TTA GCC AAT GGA GTT TCC AGA     1084
Gln Asp Gln Leu Ile Val Phe Met Ala Leu Ala Asn Gly Val Ser Arg
                300                 305                 310

ATA AAA ACA GGA CCA GTT ACA CTC CAT ACG CAA ACC GCG ATA CAT TTT     1132
Ile Lys Thr Gly Pro Val Thr Leu His Thr Gln Thr Ala Ile His Phe
            315                 320                 325

GCT GAA CAA ATA GCA AAG GCT AAA TTT ATT GTG AAG AAA TCA GAA GAT     1180
Ala Glu Gln Ile Ala Lys Ala Lys Phe Ile Val Lys Lys Ser Glu Asp
        330                 335                 340

GAA GAA GAC GCC GCT AAA GAT ACT TAT ATT ATT GAA TGC CAA GGA ATT     1228
Glu Glu Asp Ala Ala Lys Asp Thr Tyr Ile Ile Glu Cys Gln Gly Ile
    345                 350                 355

GGG ATG ACA AAT CCA AAT CTA TAG AGTATTTGCC TCTTAAATGA TACCTCATTG     1282
Gly Met Thr Asn Pro Asn Leu  *
360                 365

ATATATTGCA CTATTTCATA AATACTATAA AATAATGACT AGGAAGTAAC TTATTAAAGG     1342

CTATGACTTA AATTTGAAGA TGAAGTACAG TGTTCTAGGT TTGCTGAGAA GGCTTCATTA     1402

AATTAATCTC ACTTTGAATA TCTCCTGAGA GATGGACAAT GAAATATCAG TTGGTGGATA     1462

TGTGTGATAG CTGATTTCAA TATTGAAGTA TTGAAATAAA ATATTCTTTA CACCTGAGAA     1522

AAAAAAAA                                                             1530
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 367 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Ala Gly Pro Trp Val Glu Val Asp Gly Ser Ile Met Glu Gly Gly
 1               5                  10                  15

Gly Gln Ile Leu Arg Val Ser Thr Ala Leu Ser Cys Leu Leu Gly Leu
            20                  25                  30

Pro Leu Arg Val Gln Lys Ile Arg Ala Gly Arg Ser Thr Pro Gly Leu
        35                  40                  45

Arg Pro Gln His Leu Ser Gly Leu Glu Met Ile Arg Asp Leu Cys Asp
    50                  55                  60

Gly Gln Leu Glu Gly Ala Glu Ile Gly Ser Thr Glu Ile Thr Phe Thr
65                  70                  75                  80

Pro Glu Lys Ile Lys Gly Gly Ile His Thr Ala Asp Thr Lys Thr Ala
                85                  90                  95

Gly Ser Val Cys Leu Leu Met Gln Val Ser Met Pro Cys Val Leu Phe
            100                 105                 110

Ala Ala Ser Pro Ser Glu Leu His Leu Lys Gly Gly Thr Asn Ala Glu
        115                 120                 125

Met Ala Pro Gln Ile Asp Tyr Thr Val Met Val Phe Lys Pro Ile Val
    130                 135                 140

Glu Lys Phe Gly Phe Ile Phe Asn Cys Asp Ile Lys Thr Arg Gly Tyr
145                 150                 155                 160

Tyr Pro Lys Gly Gly Gly Glu Val Ile Val Arg Met Ser Pro Val Lys
                165                 170                 175

Gln Leu Asn Pro Ile Asn Leu Thr Glu Arg Gly Cys Val Thr Lys Ile
            180                 185                 190
```

Tyr Gly Arg Ala Phe Val Ala Gly Val Leu Pro Phe Lys Val Ala Lys
195 200 205

Asp Met Ala Ala Ala Ala Val Arg Cys Ile Arg Lys Glu Ile Arg Asp
210 215 220

Leu Tyr Val Asn Ile Gln Pro Val Gln Glu Pro Lys Asp Gln Ala Phe
225 230 235 240

Gly Asn Gly Asn Gly Ile Ile Ile Ile Ala Glu Thr Ser Thr Gly Cys
245 250 255

Leu Phe Ala Gly Ser Ser Leu Gly Lys Arg Gly Val Asn Ala Asp Lys
260 265 270

Val Gly Ile Glu Ala Ala Glu Met Leu Leu Ala Asn Leu Arg His Gly
275 280 285

Gly Thr Val Asp Glu Tyr Leu Gln Asp Gln Leu Ile Val Phe Met Ala
290 295 300

Leu Ala Asn Gly Val Ser Arg Ile Lys Thr Gly Pro Val Thr Leu His
305 310 315 320

Thr Gln Thr Ala Ile His Phe Ala Glu Gln Ile Ala Lys Ala Lys Phe
325 330 335

Ile Val Lys Lys Ser Glu Asp Glu Glu Asp Ala Ala Lys Asp Thr Tyr
340 345 350

Ile Ile Glu Cys Gln Gly Ile Gly Met Thr Asn Pro Asn Leu *
355 360 365

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1680 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Saccharomyces cerevisiae

(vii) IMMEDIATE SOURCE:
(B) CLONE: RNA cyclase

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:301..1404

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:301..1404

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
CTTTAAAAGA TATAACTTTA TAATTATAAC CCCGCGCAAA GCGTATTAAT ACTTTAAATG      60

GGTAGTCAAG TTGCTTTTAT TACTTGCACT TACTTTGAAG GAGGATGTCA GAATTACCCT     120

CGCCATATTG GTAATGTAAT TTTTCACGGA CTCGAGATTG AAAAAATTTC AGAAAAATTT     180

GCAAAAAAGT GCCGATGAGA TGAGCCTTCA TAAGCTACGT ATGTTGTAAG GGTATAACAT     240

CTGTCAAGCT TCACCATTGT CAACACTTTT CCACTAGGAT AATCCTCCAA AAGTATAGCA     300

ATG TCA TCT TCC GCC CCC AAA TAC ACC ACT TTC CAA GGG TCA CAA AAT        348
Met Ser Ser Ser Ala Pro Lys Tyr Thr Thr Phe Gln Gly Ser Gln Asn
 1               5                  10                  15

TTT AGG TTA CGG ATC GTC TTG GCA ACA TTA TCA GGG AAA CCA ATA AAA       396
Phe Arg Leu Arg Ile Val Leu Ala Thr Leu Ser Gly Lys Pro Ile Lys
            20                  25                  30

ATT GAA AAA ATC CGT TCA GGC GAC TTA AAT CCC GGT CTG AAA GAT TAT       444
Ile Glu Lys Ile Arg Ser Gly Asp Leu Asn Pro Gly Leu Lys Asp Tyr
        35                  40                  45
```

```
GAA GTG TCT TTT CTA AGG CTG ATC GAG TCG GTC ACC AAC GGG AGT GTA      492
Glu Val Ser Phe Leu Arg Leu Ile Glu Ser Val Thr Asn Gly Ser Val
 50                  55                  60

ATT GAA ATT TCT TAT ACC GGT ACT ACT GTG ATT TAT AGG CCC GGT ATC      540
Ile Glu Ile Ser Tyr Thr Gly Thr Thr Val Ile Tyr Arg Pro Gly Ile
 65                  70                  75                  80

ATA GTT GGT GGT GCA TCA ACT CAT ATC TGC CCT AGT TCC AAA CCG GTA      588
Ile Val Gly Gly Ala Ser Thr His Ile Cys Pro Ser Ser Lys Pro Val
                 85                  90                  95

GGT TAT TTC GTC GAA CCA ATG CTA TAT TTG GCT CCA TTT TCA AAA AAG      636
Gly Tyr Phe Val Glu Pro Met Leu Tyr Leu Ala Pro Phe Ser Lys Lys
            100                 105                 110

AAA TTT TCT ATA TTA TTC AAA GGC ATA ACT GCA TCT CAC AAC GAT GCC      684
Lys Phe Ser Ile Leu Phe Lys Gly Ile Thr Ala Ser His Asn Asp Ala
        115                 120                 125

GGT ATT GAA GCT ATC AAA TGG GGA CTC ATG CCC GTT ATG GAA AAA TTT      732
Gly Ile Glu Ala Ile Lys Trp Gly Leu Met Pro Val Met Glu Lys Phe
    130                 135                 140

GGC GTT AGA GAG TGC GCC CTA CAT ACT TTG AAA AGA GGT TCA CCA CCA      780
Gly Val Arg Glu Cys Ala Leu His Thr Leu Lys Arg Gly Ser Pro Pro
145                 150                 155                 160

CTA GGA GGT GGC GAA GTG CAT TTG GTT GTT GAT TCT TTG ATT GCG CAA      828
Leu Gly Gly Gly Glu Val His Leu Val Val Asp Ser Leu Ile Ala Gln
                165                 170                 175

CCT ATA ACT ATG CAT GAA ATA GAT AGG CCC ATA ATT TCA TCG ATT ACC      876
Pro Ile Thr Met His Glu Ile Asp Arg Pro Ile Ile Ser Ser Ile Thr
            180                 185                 190

GGT GTA GCA TAC TCT ACC AGA GTA AGT CCG TCG CTT GTG AAT AGA ATG      924
Gly Val Ala Tyr Ser Thr Arg Val Ser Pro Ser Leu Val Asn Arg Met
        195                 200                 205

ATC GAT GGT GCT AAG AAG GTA TTG AAA AAT CTG CAA TGC GAA GTT AAC      972
Ile Asp Gly Ala Lys Lys Val Leu Lys Asn Leu Gln Cys Glu Val Asn
    210                 215                 220
```

```
ATA ACG GCA GAT GTC TGG AGA GGT GAA AAT TCA GGG AAG AGT CCA GGC      1020
Ile Thr Ala Asp Val Trp Arg Gly Glu Asn Ser Gly Lys Ser Pro Gly
225                 230                 235                 240

TGG GGT ATT ACT TTG GTT GCT CAG TCC AAG CAG AAA GGT TGG AGT TAT      1068
Trp Gly Ile Thr Leu Val Ala Gln Ser Lys Gln Lys Gly Trp Ser Tyr
                245                 250                 255

TTT GCA GAA GAT ATC GGT GAT GCA GGT TCT ATA CCT GAA GAA CTT GGT      1116
Phe Ala Glu Asp Ile Gly Asp Ala Gly Ser Ile Pro Glu Glu Leu Gly
            260                 265                 270

GAA AAA GTT GCC TGT CAA TTA TTA GAA GAA ATA TCA AAG AGT GCA GCA      1164
Glu Lys Val Ala Cys Gln Leu Leu Glu Glu Ile Ser Lys Ser Ala Ala
        275                 280                 285

GTT GGT AGA AAC CAG CTT CCA TTA GCA ATT GTT TAC ATG GTC ATC GGG      1212
Val Gly Arg Asn Gln Leu Pro Leu Ala Ile Val Tyr Met Val Ile Gly
    290                 295                 300

AAA GAA GAT ATC GGC AGA TTG AGA ATT AAT AAG GAA CAG ATA GAC GAA      1260
Lys Glu Asp Ile Gly Arg Leu Arg Ile Asn Lys Glu Gln Ile Asp Glu
305                 310                 315                 320

AGA TTC ATA ATC CTC TTG AGA GAT ATT AAG AAG ATC TTT AAT ACT GAA      1308
Arg Phe Ile Ile Leu Leu Arg Asp Ile Lys Lys Ile Phe Asn Thr Glu
                325                 330                 335

GTC TTT TTA AAA CCA GTT GAC GAG GCG GAT AAT GAA GAC ATG ATA GCT      1356
Val Phe Leu Lys Pro Val Asp Glu Ala Asp Asn Glu Asp Met Ile Ala
            340                 345                 350

ACT ATC AAG GGT ATT GGT TTC ACA AAC ACA AGC AAA AAG ATT GCA TAG      1404
Thr Ile Lys Gly Ile Gly Phe Thr Asn Thr Ser Lys Lys Ile Ala  *
        355                 360                 365

CATGTTCTAA TATTATTCAA TAGATATTAG ACCACTGTAT ATAAGTTATA TATATATATA    1464

TGATTATTTA GGTGTAACTT CACGGACAAC TGTCTAGAGT ATAAGATTAC TGGAGGATCA    1524

TTGATGATTT CCTTTAACTC TTTTGGGTCA TCTAAAATAT CATCAAAGAG CGTATGCGAA    1584
```

AAGCGCGTTA AGAATAGATT ACTTCATGAC AAAATATAAT TTTCCAGCTC AAAAATGTTT 1644

GTTTTTTTAT GTAGACACTA TTTTCAAACT ATCTTT 1680

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 368 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Ser Ser Ser Ala Pro Lys Tyr Thr Thr Phe Gln Gly Ser Gln Asn
1 5 10 15

Phe Arg Leu Arg Ile Val Leu Ala Thr Leu Ser Gly Lys Pro Ile Lys
20 25 30

Ile Glu Lys Ile Arg Ser Gly Asp Leu Asn Pro Gly Leu Lys Asp Tyr
35 40 45

Glu Val Ser Phe Leu Arg Leu Ile Glu Ser Val Thr Asn Gly Ser Val
50 55 60

Ile Glu Ile Ser Tyr Thr Gly Thr Thr Val Ile Tyr Arg Pro Gly Ile
65 70 75 80

Ile Val Gly Gly Ala Ser Thr His Ile Cys Pro Ser Ser Lys Pro Val
85 90 95

Gly Tyr Phe Val Glu Pro Met Leu Tyr Leu Ala Pro Phe Ser Lys Lys
100 105 110

Lys Phe Ser Ile Leu Phe Lys Gly Ile Thr Ala Ser His Asn Asp Ala
115 120 125

Gly Ile Glu Ala Ile Lys Trp Gly Leu Met Pro Val Met Glu Lys Phe
130 135 140

Gly Val Arg Glu Cys Ala Leu His Thr Leu Lys Arg Gly Ser Pro Pro
145 150 155 160

Leu Gly Gly Gly Glu Val His Leu Val Val Asp Ser Leu Ile Ala Gln
165 170 175

Pro Ile Thr Met His Glu Ile Asp Arg Pro Ile Ile Ser Ser Ile Thr
180 185 190

Gly Val Ala Tyr Ser Thr Arg Val Ser Pro Ser Leu Val Asn Arg Met
195 200 205

Ile Asp Gly Ala Lys Lys Val Leu Lys Asn Leu Gln Cys Glu Val Asn
210 215 220

Ile Thr Ala Asp Val Trp Arg Gly Glu Asn Ser Gly Lys Ser Pro Gly
225 230 235 240

Trp Gly Ile Thr Leu Val Ala Gln Ser Lys Gln Lys Gly Trp Ser Tyr
245 250 255

Phe Ala Glu Asp Ile Gly Asp Ala Gly Ser Ile Pro Glu Glu Leu Gly
260 265 270

Glu Lys Val Ala Cys Gln Leu Leu Glu Glu Ile Ser Lys Ser Ala Ala
275 280 285

Val Gly Arg Asn Gln Leu Pro Leu Ala Ile Val Tyr Met Val Ile Gly
290 295 300

Lys Glu Asp Ile Gly Arg Leu Arg Ile Asn Lys Glu Gln Ile Asp Glu
305 310 315 320

Arg Phe Ile Ile Leu Leu Arg Asp Ile Lys Lys Ile Phe Asn Thr Glu
325 330 335

Val Phe Leu Lys Pro Val Asp Glu Ala Asp Asn Glu Asp Met Ile Ala
340 345 350

Thr Ile Lys Gly Ile Gly Phe Thr Asn Thr Ser Lys Lys Ile Ala *
355 360 365

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1021 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(vii) IMMEDIATE SOURCE:
(B) CLONE: RNA cyclase

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..1020

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:1..1020

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
ATG ATG AAA AGG ATG ATT GCG CTG GAT GGC GCA CAG GGC GAA GGT GGC        48
Met Met Lys Arg Met Ile Ala Leu Asp Gly Ala Gln Gly Glu Gly Gly
 1               5                  10                  15

GGG CAG ATC CTG CGC TCG GCG CTG AGC CTG TCG ATG ATA ACC GGC CAG        96
Gly Gln Ile Leu Arg Ser Ala Leu Ser Leu Ser Met Ile Thr Gly Gln
            20                  25                  30

CCA TTT ACC ATC ACC AGC ATT CGT GCC GGG CGG GCG AAA CCG GGG CTG       144
Pro Phe Thr Ile Thr Ser Ile Arg Ala Gly Arg Ala Lys Pro Gly Leu
        35                  40                  45
```

```
TTG CGC CAG CAT CTG ACC GCG GTA AAG GCG GCG ACG GAA ATT TGC GGA      192
Leu Arg Gln His Leu Thr Ala Val Lys Ala Ala Thr Glu Ile Cys Gly
     50                  55                  60

GCA ACC GTG GAA GGC GCG GAG CTG GGG TCG CAG CGT CTG CTC TTC CGG      240
Ala Thr Val Glu Gly Ala Glu Leu Gly Ser Gln Arg Leu Leu Phe Arg
 65                  70                  75                  80

CCC GGC ACC GTG CGC GGC GGC GAT TAC CGC TTT GCT ATC GGT AGC GCC      288
Pro Gly Thr Val Arg Gly Gly Asp Tyr Arg Phe Ala Ile Gly Ser Ala
                 85                  90                  95

GGA AGT TGT ACG CTG GTG CTG CAA ACG GTG CTG CCC GCG CTG TGG TTT      336
Gly Ser Cys Thr Leu Val Leu Gln Thr Val Leu Pro Ala Leu Trp Phe
            100                 105                 110

GCC GAT GGA CCT TCG CGT GTT GAA GTG AGC GGC GGC ACC GAT AAC CCG      384
Ala Asp Gly Pro Ser Arg Val Glu Val Ser Gly Gly Thr Asp Asn Pro
        115                 120                 125

CTC GGC GCC GCC TCG GAT TTT ATC CGC CGG GTG CTG GAG CCG CTG CTG      432
Leu Gly Ala Ala Ser Asp Phe Ile Arg Arg Val Leu Glu Pro Leu Leu
    130                 135                 140

GCG AAA ATA GGA ATT CAT CAG CAA ACC ACG CTG TTA CGC CAC GGT TTT      480
Ala Lys Ile Gly Ile His Gln Gln Thr Thr Leu Leu Arg His Gly Phe
145                 150                 155                 160

TAC CCT GCC GGA GGC GGT GTG GTG GCA ACG GAA GTC TCG CCC GTG GCA      528
Tyr Pro Ala Gly Gly Gly Val Val Ala Thr Glu Val Ser Pro Val Ala
                165                 170                 175

TCG TTT AAC ACC TTG CAA CTT GGC GAG CGC GGG AAC ATT GTG CAG ATG      576
Ser Phe Asn Thr Leu Gln Leu Gly Glu Arg Gly Asn Ile Val Gln Met
            180                 185                 190

CGT GGA GAA GTT CTA TTA GCT GGT GTG CCG CGC CAT GTT GCT GAG CGT      624
Arg Gly Glu Val Leu Leu Ala Gly Val Pro Arg His Val Ala Glu Arg
        195                 200                 205

GAA ATC GCT ACA CTG GCG GGG AGT TTT TCC CTG CAT GAA CAG AAT ATT      672
Glu Ile Ala Thr Leu Ala Gly Ser Phe Ser Leu His Glu Gln Asn Ile
    210                 215                 220
```

```
CAT AAC CTG CCG CGC GAC CAG GGG CCG GGT AAT ACC GTC TCG CTT GAA      720
His Asn Leu Pro Arg Asp Gln Gly Pro Gly Asn Thr Val Ser Leu Glu
225                 230                 235                 240

GTC GAA AGT GAA AAT ATC ACC GAA CGC TTT TTT GTC GTC GGT GAA AAG      768
Val Glu Ser Glu Asn Ile Thr Glu Arg Phe Phe Val Val Gly Glu Lys
                245                 250                 255

CGC GTC AGT GCC GAG GTG GTT GCG GCA CAG TTG GTG AAA GAG GTG AAA      816
Arg Val Ser Ala Glu Val Val Ala Ala Gln Leu Val Lys Glu Val Lys
            260                 265                 270

CGC TAC CTG GCA AGC ACG GCG GCG GTG GGG GAA TAT CTT GCT GAC CAA      864
Arg Tyr Leu Ala Ser Thr Ala Ala Val Gly Glu Tyr Leu Ala Asp Gln
        275                 280                 285

CTG GTG CTA CCG ATG GCG CTG GCG GGC GCG GGG GAA TTT ACG GTC GCC      912
Leu Val Leu Pro Met Ala Leu Ala Gly Ala Gly Glu Phe Thr Val Ala
    290                 295                 300

CAT CCC TCA TGC CAT CTG CTG ACC AAT ATC GCG GTG GTG GAG CGT TTC      960
His Pro Ser Cys His Leu Leu Thr Asn Ile Ala Val Val Glu Arg Phe
305                 310                 315                 320

TTG CCG GTG CGG TTT AGT TTG ATA GAA ACA GAT GGC GTA ACG CGG GTG     1008
Leu Pro Val Arg Phe Ser Leu Ile Glu Thr Asp Gly Val Thr Arg Val
                325                 330                 335

AGC ATT GAA TGA T                                                  1021
Ser Ile Glu  *
            340
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Met Lys Arg Met Ile Ala Leu Asp Gly Ala Gln Gly Glu Gly Gly
1 5 10 15

Gly Gln Ile Leu Arg Ser Ala Leu Ser Leu Ser Met Ile Thr Gly Gln
20 25 30

Pro Phe Thr Ile Thr Ser Ile Arg Ala Gly Arg Ala Lys Pro Gly Leu
35 40 45

Leu Arg Gln His Leu Thr Ala Val Lys Ala Ala Thr Glu Ile Cys Gly
50 55 60

Ala Thr Val Glu Gly Ala Glu Leu Gly Ser Gln Arg Leu Leu Phe Arg
65 70 75 80

Pro Gly Thr Val Arg Gly Gly Asp Tyr Arg Phe Ala Ile Gly Ser Ala
85 90 95

Gly Ser Cys Thr Leu Val Leu Gln Thr Val Leu Pro Ala Leu Trp Phe
100 105 110

Ala Asp Gly Pro Ser Arg Val Glu Val Ser Gly Gly Thr Asp Asn Pro
115 120 125

Leu Gly Ala Ala Ser Asp Phe Ile Arg Arg Val Leu Glu Pro Leu Leu
130 135 140

Ala Lys Ile Gly Ile His Gln Gln Thr Thr Leu Leu Arg His Gly Phe
145 150 155 160

Tyr Pro Ala Gly Gly Gly Val Val Ala Thr Glu Val Ser Pro Val Ala
165 170 175

Ser Phe Asn Thr Leu Gln Leu Gly Glu Arg Gly Asn Ile Val Gln Met
180 185 190

Arg Gly Glu Val Leu Leu Ala Gly Val Pro Arg His Val Ala Glu Arg
195 200 205

Glu Ile Ala Thr Leu Ala Gly Ser Phe Ser Leu His Glu Gln Asn Ile
210 215 220

His Asn Leu Pro Arg Asp Gln Gly Pro Gly Asn Thr Val Ser Leu Glu
225 230 235 240

Val Glu Ser Glu Asn Ile Thr Glu Arg Phe Phe Val Val Gly Glu Lys
245 250 255

Arg Val Ser Ala Glu Val Val Ala Ala Gln Leu Val Lys Glu Val Lys
260 265 270

Arg Tyr Leu Ala Ser Thr Ala Ala Val Gly Glu Tyr Leu Ala Asp Gln
275 280 285

Leu Val Leu Pro Met Ala Leu Ala Gly Ala Gly Glu Phe Thr Val Ala
290 295 300

His Pro Ser Cys His Leu Leu Thr Asn Ile Ala Val Val Glu Arg Phe
305 310 315 320

Leu Pro Val Arg Phe Ser Leu Ile Glu Thr Asp Gly Val Thr Arg Val
325 330 335

Ser Ile Glu *
340

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1092 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Saccharomyces pombe

(vii) IMMEDIATE SOURCE:
(B) CLONE: RNA cyclase

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..1092

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:1..1092

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
ATG TCA ACC GGT CAG CTC AAA CGC TTT AAA GGA TGT GAG TAT TTG ACG      48
Met Ser Thr Gly Gln Leu Lys Arg Phe Lys Gly Cys Glu Tyr Leu Thr
 1               5                  10                  15

CAT CGG TTA GTC CTT GCA ACT CTC TCG GGA ACT CCG ATT CGA GTT GAA      96
His Arg Leu Val Leu Ala Thr Leu Ser Gly Thr Pro Ile Arg Val Glu
            20                  25                  30

GGA ATT TAC CCA GAT GAA GCC GAT CCT GGT GTC AAA GAT TAC CAG GTT     144
Gly Ile Tyr Pro Asp Glu Ala Asp Pro Gly Val Lys Asp Tyr Gln Val
        35                  40                  45

TCG TTT TTA AGG CTG TTG GAG AAG TTG ACC AAT GGC AGT GTC ATA GAA     192
Ser Phe Leu Arg Leu Leu Glu Lys Leu Thr Asn Gly Ser Val Ile Glu
    50                  55                  60

ATT TCT TAC ACA GGT ACT TCG TTC ATA TAT CGC CCA GGA AAT ATT ATA     240
Ile Ser Tyr Thr Gly Thr Ser Phe Ile Tyr Arg Pro Gly Asn Ile Ile
65                  70                  75                  80

GGA GGT CGT GTA GTC CAT GAT TGT CCA ACT ACC AAA GGG ATA GGC TAT     288
Gly Gly Arg Val Val His Asp Cys Pro Thr Thr Lys Gly Ile Gly Tyr
                85                  90                  95

TTT TTG GAA CCA ATT CTA ATA CTT TGT CTT TTT GCA AAA ACC CCT ACT     336
Phe Leu Glu Pro Ile Leu Ile Leu Cys Leu Phe Ala Lys Thr Pro Thr
            100                 105                 110

TCA CTT ACT CTT ACT GGA GTT ACT TCC AGT AAT GAA GAT ATT GGT GTC     384
Ser Leu Thr Leu Thr Gly Val Thr Ser Ser Asn Glu Asp Ile Gly Val
        115                 120                 125
```

GAT GTT TTG CGC ACT AGT GTA TTA CCC TCT CTT CAA AAG CGT TTT CAA 432
Asp Val Leu Arg Thr Ser Val Leu Pro Ser Leu Gln Lys Arg Phe Gln
130 135 140

GTT GGT GAT GAA CTT GAG TTG CGC ATA TTA AAG CGT GGT TCT GCT CCT 480
Val Gly Asp Glu Leu Glu Leu Arg Ile Leu Lys Arg Gly Ser Ala Pro
145 150 155 160

GGC GGA GGA GGT GAA GTG AAT TTT TTA TGT CCT GTA ATT AAA GAG TCT 528
Gly Gly Gly Gly Glu Val Asn Phe Leu Cys Pro Val Ile Lys Glu Ser
165 170 175

TTA CCT CCC ATT CGT CTC TCA GAA TTT GGA AGA GTC TTC CGA ATC AGA 576
Leu Pro Pro Ile Arg Leu Ser Glu Phe Gly Arg Val Phe Arg Ile Arg
180 185 190

GGA ATT GCT TCT TCA ACT AGA GTA TCT CCC GCT TTT GCA AAT CGC TTA 624
Gly Ile Ala Ser Ser Thr Arg Val Ser Pro Ala Phe Ala Asn Arg Leu
195 200 205

GTA GAA TCT GCT CGC GGA GTT TTA AAT CCT TTT ATA CCC GAT GTT TTC 672
Val Glu Ser Ala Arg Gly Val Leu Asn Pro Phe Ile Pro Asp Val Phe
210 215 220

ATT TAT ACA GAC GTT CGC CGT GGA GAT GAA TGT GGT AAC AGT CCT GGG 720
Ile Tyr Thr Asp Val Arg Arg Gly Asp Glu Cys Gly Asn Ser Pro Gly
225 230 235 240

TAT TCT ATC ACT TTA GTT GCA GAA ACT AAT AAG GGT TGC TCT TAT GCT 768
Tyr Ser Ile Thr Leu Val Ala Glu Thr Asn Lys Gly Cys Ser Tyr Ala
245 250 255

GCC GAG CAT TGT GGT GAA GCT GGA GAA ACA CCA GAA GAT GTT GGT TCC 816
Ala Glu His Cys Gly Glu Ala Gly Glu Thr Pro Glu Asp Val Gly Ser
260 265 270

TTT TGC GCA AAG AAA TTA TTG GAA GTT ATT GAA TCT GGT GGT TGT GTA 864
Phe Cys Ala Lys Lys Leu Leu Glu Val Ile Glu Ser Gly Gly Cys Val
275 280 285

```
GAT CCA TAT ACA CAA CCT TCA ACT TTA ACC GGG ATG CTA CTA TCC TCT      912
Asp Pro Tyr Thr Gln Pro Ser Thr Leu Thr Gly Met Leu Leu Ser Ser
    290                 295                 300

GAA GAC GTT AAC ACA ATA GTT GTG GGT CAA CTT GGT ATT ACT TCG CAA      960
Glu Asp Val Asn Thr Ile Val Val Gly Gln Leu Gly Ile Thr Ser Gln
305                 310                 315                 320

TTA GTT GTT TTT CTT CGG GAT GTG AAG GCA CTT TTC AAT TGC GAA TAT     1008
Leu Val Val Phe Leu Arg Asp Val Lys Ala Leu Phe Asn Cys Glu Tyr
                325                 330                 335

AGA TTT AAA GAG CTC GAA TCA GGC CAG GTG GAG ATG TCA TGC TTA GGA     1056
Arg Phe Lys Glu Leu Glu Ser Gly Gln Val Glu Met Ser Cys Leu Gly
            340                 345                 350

AAA GGA TAT TTA AAT GTC AAT CGC CGC ATT CAA TAG                     1092
Lys Gly Tyr Leu Asn Val Asn Arg Arg Ile Gln  *
        355                 360
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 364 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Met Ser Thr Gly Gln Leu Lys Arg Phe Lys Gly Cys Glu Tyr Leu Thr
  1               5                  10                  15

His Arg Leu Val Leu Ala Thr Leu Ser Gly Thr Pro Ile Arg Val Glu
            20                  25                  30

Gly Ile Tyr Pro Asp Glu Ala Asp Pro Gly Val Lys Asp Tyr Gln Val
        35                  40                  45

Ser Phe Leu Arg Leu Leu Glu Lys Leu Thr Asn Gly Ser Val Ile Glu
    50                  55                  60
```

Ile Ser Tyr Thr Gly Thr Ser Phe Ile Tyr Arg Pro Gly Asn Ile Ile
65 70 75 80

Gly Gly Arg Val Val His Asp Cys Pro Thr Thr Lys Gly Ile Gly Tyr
85 90 95

Phe Leu Glu Pro Ile Leu Ile Leu Cys Leu Phe Ala Lys Thr Pro Thr
100 105 110

Ser Leu Thr Leu Thr Gly Val Thr Ser Ser Asn Glu Asp Ile Gly Val
115 120 125

Asp Val Leu Arg Thr Ser Val Leu Pro Ser Leu Gln Lys Arg Phe Gln
130 135 140

Val Gly Asp Glu Leu Glu Leu Arg Ile Leu Lys Arg Gly Ser Ala Pro
145 150 155 160

Gly Gly Gly Gly Glu Val Asn Phe Leu Cys Pro Val Ile Lys Glu Ser
165 170 175

Leu Pro Pro Ile Arg Leu Ser Glu Phe Gly Arg Val Phe Arg Ile Arg
180 185 190

Gly Ile Ala Ser Ser Thr Arg Val Ser Pro Ala Phe Ala Asn Arg Leu
195 200 205

Val Glu Ser Ala Arg Gly Val Leu Asn Pro Phe Ile Pro Asp Val Phe
210 215 220

Ile Tyr Thr Asp Val Arg Arg Gly Asp Glu Cys Gly Asn Ser Pro Gly
225 230 235 240

Tyr Ser Ile Thr Leu Val Ala Glu Thr Asn Lys Gly Cys Ser Tyr Ala
245 250 255

Ala Glu His Cys Gly Glu Ala Gly Glu Thr Pro Glu Asp Val Gly Ser
260 265 270

Phe Cys Ala Lys Lys Leu Leu Glu Val Ile Glu Ser Gly Gly Cys Val
275 280 285

Asp Pro Tyr Thr Gln Pro Ser Thr Leu Thr Gly Met Leu Leu Ser Ser
290 295 300

Glu Asp Val Asn Thr Ile Val Val Gly Gln Leu Gly Ile Thr Ser Gln
305 310 315 320

Leu Val Val Phe Leu Arg Asp Val Lys Ala Leu Phe Asn Cys Glu Tyr
325 330 335

Arg Phe Lys Glu Leu Glu Ser Gly Gln Val Glu Met Ser Cys Leu Gly
340 345 350

Lys Gly Tyr Leu Asn Val Asn Arg Arg Ile Gln *
355 360

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 518 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(C) INDIVIDUAL ISOLATE: HeLa

(vii) IMMEDIATE SOURCE:
(B) CLONE: RNA cyclase 2

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:2..517

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:2..517

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TCTACGAGGA GTGACCAATG ATCAGGTTGA CCCTTCAGTT GATGTTCTTA AGGCAACAGC 60

ACTCCCTTTG TTGAAACAAT TTGGGATTGA TGGTGAATCA TTTGAACTGA AGATTGTGCG 120

ACGGGGAATG CCTCCCGGAG GAGGAGGCGA AGTGGTTTTC TCATGTCCTG TGAGGAAGGT 180

CTTGAAGCCC ATTCAACTCA CAGATCCAGG AAAAATCAAA CGTATTANAG GAATGGCGTA 240

CTCTGTACGT GTGTCACCTC AGATGGCGAA CCGGATTGTG GATTCTGCAA GGAGCATCCT 300

CAACAAGTTC ATACCTGATA TCTATATTTA CACAGATCAC ATGAAAGGAG TCAACTCTGG 360

GAAGTCTCCG GGCTTTGGGT TGTCACTGGT TGCTGAGACC ACCAGTGGCA CCTTCCTCAG 420

TGCTGAACTG GCCTCCAACC CCCANGGCCA GGGAACAGCA GTACTTCCAG AAGACCTTGG 480

CAGGAACTGT GCCCGGCTGC TGCTGGAAGA AATCTACC 518

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 708 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Drosphila melanogaster

(vii) IMMEDIATE SOURCE:
(B) CLONE: RNA cyclase

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..708

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:1..708

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
GTC GAC GAG GGC CTC GAA CTC AAG GTT GTG CGC CGA GGA GTT GCT CCG      48
Val Asp Glu Gly Leu Glu Leu Lys Val Val Arg Arg Gly Val Ala Pro
 1               5                  10                  15

CTG GGC GGC GGA GAG ATC ATC TTC AGG TGT CCG GTG CGA AAG AGC CTG      96
Leu Gly Gly Gly Glu Ile Ile Phe Arg Cys Pro Val Arg Lys Ser Leu
            20                  25                  30

CGT GCC ATA CAA TTT CAG TCA CAG GGC ATG GTT AAA CGC ATC CGG GGC     144
Arg Ala Ile Gln Phe Gln Ser Gln Gly Met Val Lys Arg Ile Arg Gly
        35                  40                  45

ACT GTT TAT GCC TGC AAA GTC TCT CCG GCG ATG GCC AAT CGC ACC GTG     192
Thr Val Tyr Ala Cys Lys Val Ser Pro Ala Met Ala Asn Arg Thr Val
    50                  55                  60

GAG GCG GCT AAG GGA TGC ATG CTA AAG TTC CTG CCG GAT GTC TAT ATT     240
Glu Ala Ala Lys Gly Cys Met Leu Lys Phe Leu Pro Asp Val Tyr Ile
65                  70                  75                  80

TAC ACA GAT CAG AAC AAA GGC AAG ATG TCG GGT AAT TCG CCC GGA TTT     288
Tyr Thr Asp Gln Asn Lys Gly Lys Met Ser Gly Asn Ser Pro Gly Phe
                85                  90                  95

GGC ATC TGT CTG ATT GCC GAG ACT ACG GAC GGT GTG TGC TTC GCC GCC     336
Gly Ile Cys Leu Ile Ala Glu Thr Thr Asp Gly Val Cys Phe Ala Ala
            100                 105                 110

GAT TGC TGT TCC AAC ACA AGG GAG GAG TCG GAG GAT ACA CCA TCC ATA     384
Asp Cys Cys Ser Asn Thr Arg Glu Glu Ser Glu Asp Thr Pro Ser Ile
        115                 120                 125
```

```
CCC GAG AAT CTG GGC AAG GAG GTG GCA CTG CGT TTG CTA GAC GAG ATT      432
Pro Glu Asn Leu Gly Lys Glu Val Ala Leu Arg Leu Leu Asp Glu Ile
    130                 135                 140

TAT CGC GGT GGC TGT GTA GAC TCC AGC TAT CAA TGG CTG GCG GCA CTC      480
Tyr Arg Gly Gly Cys Val Asp Ser Ser Tyr Gln Trp Leu Ala Ala Leu
145                 150                 155                 160

TAT ATA GCA CTG GGG CAA AAG CAT GTC TCC AAA TTC CTC ACC GGT GCC      528
Tyr Ile Ala Leu Gly Gln Lys His Val Ser Lys Phe Leu Thr Gly Ala
                165                 170                 175

CTG TCC AAT TAC ACC GTT CAC TTT CTG CAA CAT TTG CGC GAC TTC TTC      576
Leu Ser Asn Tyr Thr Val His Phe Leu Gln His Leu Arg Asp Phe Phe
            180                 185                 190

TCG ATC ACC TTC AAG CTG GAG AAT CCC GAG GCG GAG GAC GAG GAC GAA      624
Ser Ile Thr Phe Lys Leu Glu Asn Pro Glu Ala Glu Asp Glu Asp Glu
        195                 200                 205

GCG GAG AAT GTG CGT GGT GCC CAG AAG GTT CTT ATG GCA TGT GTT GGA      672
Ala Glu Asn Val Arg Gly Ala Gln Lys Val Leu Met Ala Cys Val Gly
    210                 215                 220

ATT GGC TAC ACA AAT ATA AAT AAG CGC GTT ATA TAA                      708
Ile Gly Tyr Thr Asn Ile Asn Lys Arg Val Ile  *
225                 230                 235
```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 236 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
Val Asp Glu Gly Leu Glu Leu Lys Val Val Arg Arg Gly Val Ala Pro
  1               5                  10                  15
```

Leu Gly Gly Gly Glu Ile Ile Phe Arg Cys Pro Val Arg Lys Ser Leu
20 25 30

Arg Ala Ile Gln Phe Gln Ser Gln Gly Met Val Lys Arg Ile Arg Gly
35 40 45

Thr Val Tyr Ala Cys Lys Val Ser Pro Ala Met Ala Asn Arg Thr Val
50 55 60

Glu Ala Ala Lys Gly Cys Met Leu Lys Phe Leu Pro Asp Val Tyr Ile
65 70 75 80

Tyr Thr Asp Gln Asn Lys Gly Lys Met Ser Gly Asn Ser Pro Gly Phe
85 90 95

Gly Ile Cys Leu Ile Ala Glu Thr Thr Asp Gly Val Cys Phe Ala Ala
100 105 110

Asp Cys Cys Ser Asn Thr Arg Glu Glu Ser Glu Asp Thr Pro Ser Ile
115 120 125

Pro Glu Asn Leu Gly Lys Glu Val Ala Leu Arg Leu Leu Asp Glu Ile
130 135 140

Tyr Arg Gly Gly Cys Val Asp Ser Ser Tyr Gln Trp Leu Ala Ala Leu
145 150 155 160

Tyr Ile Ala Leu Gly Gln Lys His Val Ser Lys Phe Leu Thr Gly Ala
165 170 175

Leu Ser Asn Tyr Thr Val His Phe Leu Gln His Leu Arg Asp Phe Phe
180 185 190

Ser Ile Thr Phe Lys Leu Glu Asn Pro Glu Ala Glu Asp Glu Asp Glu
195 200 205

Ala Glu Asn Val Arg Gly Ala Gln Lys Val Leu Met Ala Cys Val Gly
210 215 220

Ile Gly Tyr Thr Asn Ile Asn Lys Arg Val Ile *
225 230 235

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

CACATGCGTG AATATCGAC 19

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "snthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TCYTGSACDG GYTGRATRTT 20

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GAYGGNTCHA THATGGARGG 20

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GGCTGRATRT TVACRTASAG RTC 23

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic oligonucleotide"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CCCCACCCCG 10

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic oligonucleotide"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AAAATAAAAG 10

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GAGGAGAAAG AATTCATGGC GGGGCCGTGG 30

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

AGTGGTGATG GTGGAATTCC TATAGATTTG 30

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TGCGTAAAAG GATCCATGGT GAAAAGGATG 30

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CTTACATCTC TGGATCCTTC AATGCTCAC 29

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

ATCTCTAGAG ATACCTGTTG CTGCTTAATC 30

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGTCGCGGAT CCCTCATGCC ATCTGCTGAC 30

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GGTCGCGGAT CCCTCATGCC ATCTGCTGAC 30

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

CTTACTTTGT CGACCTGGCA CAAAAAGAGA TG 32

What is claimed is:

1. A method for producing RNA 3'-terminal phosphate cyclase or a derivative thereof comprising expressing a nucleic acid sequence encoding said cyclase in a recombinant host cell, comprising the steps of transfecting a host cell with a vector comprising a nucleic acid sequence encoding RNA 3'-terminal phosphate cyclase, culturing the host cell to express RNA 3'-terminal phosphate cyclase and purifying RNA 3'-terminal phosphate cyclase from the cell culture.

2. A method according to claim 1, wherein the nucleic acid sequence is a sequence capable of hybridising with any one of the sequences set forth in SEQ ID No. 1, 3, 5, 7, 9 or 11.

3. A method according to claim 1 or claim 2, wherein the RNA 3'-terminal phosphate cyclase has homology to SEQ ID No. 2, 4, 6, 8 or 10.

4. A method according to claim 3, wherein the RNA 3'-terminal phosphate cyclase is human RNA 3'-terminal phosphate cyclase (SEQ ID No. 2 or SEQ ID No. 10), S. cerevisiae RNA 3'-terminal phosphate cyclase (SEQ ID No. 4), E. coli RNA 3'-terminal phosphate cyclase (SEQ ID No. 6), S. pombe RNA 3'-terminal phosphate cyclase (SEQ ID No. 8) or D. melanogaster RNA 3'-terminal phosphate cyclase (SEQ ID No. 12).

5. An RNA 3'-terminal phosphate cyclase in substantially pure form, producible by a method according to any one of claims 1 to 4.

6. A nucleic acid encoding an RNA 3'-terminal phosphate cyclase producible by a method according to any one of claims 1 to 4.

7. A nucleic acid according to claim 6 which is capable of hybridising with the nucleic acid sequences set forth in SEQ ID No. 1, 3, 5, 7, 9 or 11.

8. A nucleic acid according to claim 7 which is DNA and has the sequence set forth in SEQ ID No. 1, 3, 5, 7, 9 or 11.

9. An expression vector comprising a nucleic acid sequence according to any one of claims 6 to 8.

10. A host cell transformed with an expression vector according to claim 9.

11. Use of a nucleic acid according to any one of claims 6 to 8 for the production of RNA 3'-terminal phosphate cyclase.

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 97/02566

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N9/00 C12N15/52

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | WO 89 01036 A (CELLTECH LTD) 9 February 1989<br>see the whole document<br>--- | 1-11 |
| Y | VICENTE O., FILIPOWICZ W.: "Purification of RNA 3'-terminal phosphate cyclase from HeLa cells. Covalent modification of the enzyme with different nucleotides"<br>EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 176, no. 2, 1988, BERLIN, DE, pages 431-440, XP002044539<br>cited in the application<br>see the whole document<br>--- | 1-11 |
| A | EP 0 543 137 A (AMERICAN CYANAMID CO.) 26 May 1993<br>see abstract<br>see claims<br>---<br>-/-- | 1-11 |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 24 October 1997 | 11. 11. 97 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | Panzica, G |

Form PCT/ISA/210 (second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 97/02566

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X,P | GENSCHICK P. ET AL.: "The human RNA 3' -terminal phosphate cyclase is a member of a new family of proteins conserved in Eucarya, Bacteria and Archaea" EMBO JOURNAL, vol. 16, no. 10, 1997, EYNSHAM, OXFORD GB, pages 2955-2967, XP002044540 see the whole document --- | 1-11 |
| A | M. INNIS, D. GELFAND, SNINSKY, T. WHITE: "PCR Protocols. A guide to methods and applications" 1990 , ACADEMIC PRESS INC., HARCOURT BRACE JOVANOVICH, PUBLISHERS , SAN DIEGO NEW YORK BOSTON LONDON SYDNEY TOKYO TORONTO XP002044541 Chapter 6: "cDNA cloning using degenerate primers" by C.C. Lee and C.T. Caskey ----- | |

1

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Internal Application No
PCT/EP 97/02566

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 8901036 A | 09-02-89 | DE 3880468 A | 27-05-93 |
| | | EP 0323997 A | 19-07-89 |
| | | JP 2505268 B | 05-06-96 |
| | | JP 2500330 T | 08-02-90 |
| | | US 5591639 A | 07-01-97 |
| | | US 5658759 A | 19-08-97 |
| EP 0543137 A | 26-05-93 | US 5334521 A | 02-08-94 |
| | | AU 666146 B | 01-02-96 |
| | | AU 2844992 A | 20-05-93 |
| | | CA 2082986 A | 19-05-93 |
| | | JP 5276961 A | 26-10-93 |
| | | NZ 245068 A | 26-05-95 |
| | | US 5578481 A | 26-11-96 |
| | | ZA 9208877 A | 12-05-93 |